

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: September 1, 2005, 19:14:15 ; Search time 166 Seconds

(without alignment)
1311.723 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MVRPQNKRFILILNGVWN.....TRDRQPKVAHYLRRLMSRV 563

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_15Dec04:*

1: geneseqp19808:*

2: geneseqp19908:*

3: geneseqp20008:*

4: geneseqp20018:*

5: geneseqp20028:*

6: geneseqp20038:*

7: geneseqp20048:*

8: geneseqp20058:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2999	99.9	563	AA828406	Aab28406 Thermotoga
2	2999	99.9	563	ADN20342	Adn20342 Bacterial
3	932	31.1	602	AAW93820	AAW93820 Bacillus
4	929	31.0	602	AAW93825	AAW93825 Bacillus
5	929	31.0	602	AAW93822	AAW93822 Bacillus
6	929	31.0	602	AAW93826	AAW93826 Bacillus
7	929	31.0	602	AAW93822	AAW93822 Bacillus
8	929	31.0	602	AAW93822	AAW93822 Bacillus
9	929	31.0	602	AAW93822	AAW93822 Bacillus
10	920	30.8	615	AAW93821	AAW93821 Bacillus
11	920	30.7	602	AAW93821	AAW93821 Bacillus
12	905	30.6	602	AAW93821	AAW93821 Bacillus
13	902.5	30.2	656	AAW93821	AAW93821 Bacillus
14	899.5	30.0	648	AAW93821	AAW93821 Bacillus
15	897	29.9	643	AAW93821	AAW93821 Bacillus
16	894.5	29.8	648	AAW93821	AAW93821 Bacillus
17	894.5	29.4	648	AAW93821	AAW93821 Bacillus
18	882	29.3	603	AAW93827	AAW93827 E. coli
19	879	29.3	603	AAW93824	AAW93824 Human GUS
20	879	29.3	603	AAW93824	AAW93824 Human GUS
21	879	29.3	603	AAW93824	AAW93824 Human GUS
22	879	29.3	603	AAW93824	AAW93824 Human GUS
23	879	29.3	603	AAW93824	AAW93824 Human GUS
24	879	29.3	603	AAW93824	AAW93824 Human GUS
25	879	29.3	603	AAW93824	AAW93824 Human GUS

26	877	29.2	618	7	ADL01666	ADL01666 Modified
27	875.5	29.2	602	2	AAW93827	AAW93827 Beta-gluc
28	875	29.2	603	5	ABB84107	ABB84107 GUS prote
29	875	29.2	604	7	ADD27986	ADD27986 Beta-gluc
30	875	29.2	659	7	ADD27989	ADD27989 Oleosin/b
31	875	29.2	850	7	ADD27991	ADD27991 Oleosin/b
32	872.5	29.1	602	2	AAW42429	AAW42429 Escherich
33	872.5	29.1	607	8	ADS26258	ADS26258 Bacterial
34	872.5	29.1	711	6	ABR83626	ABR83626 SUMO-beta
35	872.5	29.1	1242	5	ABB81108	ABB81108 LUC-U3'-U
36	872.5	29.1	1242	6	ABB84637	ABB84637 LUC-U3'-U
37	870	29.0	832	2	AAW04302	AAW04302 Antibody/
38	869.5	29.0	602	1	AAW82948	AAW82948 Beta-gluc
39	865.5	28.8	613	2	AAW93823	AAW93823 E. coli G
40	865.5	28.8	613	3	AAW93828	AAW93828 Human GUS
41	865.5	28.8	613	3	AAW93828	AAW93828 Human GUS
42	865.5	28.8	633	4	AAW62276	AAW62276 Mutant be
43	865.5	28.8	633	4	AAW62271	AAW62271 Heavy cha
44	865.5	28.8	651	4	AAW02443	AAW02443 Human bec
45	865.5	28.8	651	7	ADD45481	ADD45481 Human Pro

ALIGNMENTS

RESULT 1
AA828406
ID AAB28406 standard; protein; 563 AA.
XX
AC AAB28406;
XX
DT 26-JAN-2001 (first entry)
XX
DE Thermotoga maritima beta-glucuronidase.
XX
KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW transgenic insect; marker; glucuronide detoxification.
XX
OS Thermotoga maritima.
XX
PN WO200055333-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US007107.
XX
PR 17-MAR-1999; 99US-00270957.
XX
PA (CAMP-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Mayer JE;
XX
DR WPI: 2000-647075/62.
XX
DR N-PSDB; AAA07937.
XX
PT Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.
XX
PS Claim 3; Fig 5B; 116pp; English.
XX
XX The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
XX genes were obtained from six different genera: Enterobacter/Salmonella,
XX Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can
XX be used as a reporter/effector molecule for transgene constructions and
XX in vitro diagnostic applications. It may also be used to generate
XX sentinel plants that serve as bioindicators of environmental status. It
XX may be used to generate transgenic insects for tracking insect
XX populations or to facilitate the development of a bioassay for compounds
XX that affect molecules critical for insect development (e.g. juvenile
XX hormone). Secreted GUS may also serve as a marker for beneficial fungi
XX destined for release into the environment. In animal systems, secreted

CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies,
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable

XX Sequence 563 AA:

Query Match 99.9%; Score 2999; DB 3; Length 563;
 Best Local Similarity 100.0%; Pred. No. 1,1e-236;
 Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRPQNRKRRFILLINGVNNLEVTSKDRPIAVPGSMNEQYODLCYEEGPFYKTTFYVPK 60
 DB 1 MVRPQNRKRRFILLINGVNNLEVTSKDRPIAVPGSMNEQYODLCYEEGPFYKTTFYVPK 60
 QY 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVYGKKSSENELRVVENRL 120
 DB 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVYGKKSSENELRVVENRL 120
 QY 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 DB 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 QY 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 DB 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 QY 181 EPEKLGKVKVKEVSEEAQVGMITKLGEEBKIRTSNRFVEGFEILENARFWSLEDPY 240
 DB 181 EPEKLGKVKVKEVSEEAQVGMITKLGEEBKIRTSNRFVEGFEILENARFWSLEDPY 240
 QY 241 LVPPLKELEKDYTLDIGRTISWDEKRLYLNGKPVFLKFGKHEBFPLGGGTFFYPLMI 300
 DB 241 LVPPLKELEKDYTLDIGRTISWDEKRLYLNGKPVFLKFGKHEBFPLGGGTFFYPLMI 300
 QY 301 KDFNLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPRVGITRHYNDETOKIAD 360
 DB 301 KDFNLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPRVGITRHYNDETOKIAD 360
 QY 361 NIRRMIDRKNHPSVTIMSVANEPESNHDAGCFALYETANENDRTPPVVMSMDAP 420
 DB 361 NIRRMIDRKNHPSVTIMSVANEPESNHDAGCFALYETANENDRTPPVVMSMDAP 420
 QY 421 DERTBDVALKYFDIYCVNRYGYIYQGRIBEGLOALEMDIEELVABRHKPIFVTEFGAD 480
 DB 421 DERTBDVALKYFDIYCVNRYGYIYQGRIBEGLOALEMDIEELVABRHKPIFVTEFGAD 480
 QY 481 AIAGIHYDPQMFSEEQALVEKTRILLLKDDYIIGTHVMAFADFKTPQNVRRPILNHK 540
 DB 481 AIAGIHYDPQMFSEEQALVEKTRILLLKDDYIIGTHVMAFADFKTPQNVRRPILNHK 540
 QY 541 GVFTDRDQPKVAHYLRRLMSEV 563
 DB 541 GVFTDRDQPKVAHYLRRLMSEV 563

RESULT 2

ID ADN20342 estandard; protein; 563 AA.

AC ADN20342;

DT 02-DEC-2004 (first entry)

DE Bacterial polypeptide #2995.

XX Recombinant DNA construct; transformed plant; improved plant property;
 XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
 XX pathogen tolerance; pest tolerance; plant disease resistance;
 XX cell cycle pathway modification; plant growth regulator;
 XX homologous recombination; seed oil yield; protein yield; carbohydrate;
 XX nitrogen; phosphorus; photocynthesis; lignin; galactomannan;

KW bacterial polypeptide.

XX Bacteria.

OS US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI: 2004-061375/06.

XX Claim 1; SEQ ID NO 2995; 122p; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 563 AA:

Query Match 99.9%; Score 2999; DB 8; Length 563;
 Best Local Similarity 99.8%; Pred. No. 1,1e-236;
 Matches 562; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVRPQNRKRRFILLINGVNNLEVTSKDRPIAVPGSMNEQYODLCYEEGPFYKTTFYVPK 60
 DB 1 MVRPQNRKRRFILLINGVNNLEVTSKDRPIAVPGSMNEQYODLCYEEGPFYKTTFYVPK 60
 QY 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVYGKKSSENELRVVENRL 120
 DB 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVYGKKSSENELRVVENRL 120
 QY 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 DB 121 KVGGFPSKVPDSGTHVGFSGFPFPPANDFPPYGGIIRPVLIETFDHARILDIWDTSES 180
 QY 181 EPEKLGKVKVKEVSEEAQVGMITKLGEEBKIRTSNRFVEGFEILENARFWSLEDPY 240
 DB 181 EPEKLGKVKVKEVSEEAQVGMITKLGEEBKIRTSNRFVEGFEILENARFWSLEDPY 240

```
Db 181 EPEKLGKVKYKIEVSEAVGOEMTIKLGEEKKIRTSNRFVEGEFLEENARFMSLEDPY 240
Qy 241 LYLKAYLEKDEXTLDIGRTISWDEKRLYLNGKPVFLKGFNGHEEPVVGOSTFYPLMI 300
Db 241 LYLKAYLEKDEXTLDIGRTISWDEKRLYLNGKPVFLKGFNGHEEPVVGOSTFYPLMI 300
Qy 301 KDFNLKMINANSFRTSHYPYSEBMDLADRLGILVDEAPHYGITRYHNPETOKIAED 360
Db 301 KDFNLKMINANSFRTSHYPYSEBMDLADRLGILVDEAPHYGITRYHNPETOKIAED 360
Qy 361 NIRMIDRKHQHPSVIMWSVANEBSNHPABEGFFKALYETANEMDRTPVVMVSMMDAP 420
Db 361 NIRMIDRKHQHPSVIMWSVANEBSNHPABEGFFKALYETANEMDRTPVVMVSMMDAP 420
Qy 421 DESTRVALKYPVICVNRYYGWTYOGRIEELGQALEKDIIEELYARHKKPIFVTEFGAD 480
Db 421 DESTRVALKYPVICVNRYYGWTYOGRIEELGQALEKDIIEELYARHKKPIFVTEFGAD 480
Qy 481 AINGIHYDPQMFSEYQALVEKTRILLLKDYIIGTHVMAFADFKTPQNVRRPILNKK 540
Db 481 AINGIHYDPQMFSEYQALVEKTRILLLKDYIIGTHVMAFADFKTPQNVRRPILNKK 540
Qy 541 GVFTRDROPKLVAVHLRLMSEV 563
Db 541 GVFTRDROPKLVAVHLRLMSEV 563

RESULT 3
AAW93820 standard; protein; 602 AA.
XX
AC AAW93820;
XX
DT 25-JUN-1999 (first entry)
XX
DE Bacillus sp. GUS protein.
XX
KM GUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW receptor molecule; diagnostic tool; transgene construction; plant;
XX insect; cleavage; detoxification; glucuronide.
XX
OS Bacillus sp.
XX
PN WO9913085-A2.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98WO-US019217.
XX
PR 09-SEP-1997; 97US-0058263P.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Kilian A, Keese PK;
XX
DR WPI; 1999-229241/19.
XX
PT New isolated microbial beta-glucuronidase.
XX
PS Claim 4; Fig 3; 76bp; English.
XX
CC This invention describes a novel secreted form of Bacillus sp. beta-
glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a
reporter/effector molecule and as a diagnostic tool. The products of the
invention can be used as markers for transgene constructions, e.g. in
plants or insects. They can also be used for the cleavage and
detoxification of glucuronides and to examine conjugation patterns of
glucuronides
CC
XX
SQ Sequence 602 AA;

Query Match 31.1%; Score 932; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.9e-67;
```

```
Matches 221; Conservative 94; Mismatches 223; Indels 82; Gaps 15;
Qy 1 MRPQNKRRKRPILLNGVNLSEV-----TSKDRPIAVGSMNE--QYDLC 44
Db 1 MRPITETRGVDFLGVNMFKLVDYKGLBEKYESKLDITISMAVSSSYNDIGVKEIR 60
Qy 45 YEEGPFYKTFPVPKLSQKHRLYFAAVNTDCEVFLNGEKGGEHIEYLPPEVDNPK 104
Db 61 NHIGVWYREFPVPAVLDQRIVLRFSGATHKAIYVNGELVHEKGGFLPEAEINNS 120
Qy 105 VSGENELRVVENRLKVGFPSPKVPDSGTHVGFPGS-----PPANFDFFPYG 154
Db 121 LRQGMNRVAVVADNLL-----DSTLPVGLYSEHBEGLGKVINKKRPFDFPNTA 170
Qy 155 GIIRPVLIEFTDARLIDMVDTSSEEPKLGKVKIEVSEAVGOEMTIKLG--EEB 212
Db 171 GHRPVKIYTFPTVEDISVTFDFNGP---TGYVTVVDFQ---GKAETVAVSVVDEE 223
Qy 213 KKIRTSNRFVEGEFLEENARFMSLEBDPYLYPLKLEKDEXTLDI-----GIRTSWDEK 267
Db 224 GKVVASTBGLSGVLEIPNVLMEPLNTLYQIKVELNDGLTIDVEEPPGVRTVAVNDG 283
Qy 268 RLYLNGKPVFLKGFNGHEEPVVGOSTFYPLMIKDFNLKMINANSFRTSHYPYSEBMD 327
Db 284 KFLINNKPFYFGFGKHEDTPINGRGFNASVMDPILKMGANSFRTAHYPSSELMR 343
Qy 328 LADRLGILVIDAPHYGITRYHN-----PETOKIA-----EDNIRMIDR 368
Db 344 LADREGLVVIDETPAVGV---HLNFMATTLGEGSERVSTWEEKIRTFEHHQVLRRLVGR 400
Qy 369 HKNHPSVIMWSVANEBSNHPABEGFFKALYETANEMD--RTRPVVMVSMMDAPDESTRV 427
Db 401 DKNHPSVIMWSVANEBSNHPABEGFFKALYETLDELDQKRPVTVKVMKTPP--TDK 458
Qy 428 ALKYPVICVNRYYGWTYOGRIEELGQALEKDIIEELYARH---KPIFVTEFGADAIA 483
Db 459 VAEILDVIALNKNYNGWIPDGDLE---AAKVRLOEFHAMNRCGKRPIMITEYAGDITVA 515
Qy 484 GIHYDPQMFSEYQALVEKTRILLLKDYIIGTHVMAFADFKTPQNVRRPILNKKYF 543
Db 516 GFHDIDPVWFTREYQVEYQANHVVDFEENFVGQAMFADPATSQGVARVQGNKKGVF 575
Qy 544 TRDROKLVAVHLRLMSEV 563
Db 576 TRDRKPKLAHAVFRERWTNI 595

RESULT 4
AAW93825
ID AAW93825 standard; protein; 602 AA.
XX
AC AAW93825;
XX
DT 25-JUN-1999 (first entry)
XX
DE Bacillus sp. codon optimised GUS protein.
XX
KM GUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW receptor molecule; diagnostic tool; transgene construction; plant;
XX insect; cleavage; detoxification; glucuronide.
XX
OS Bacillus sp.
XX
PN WO9913085-A2.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98WO-US019217.
XX
PR 09-SEP-1997; 97US-0058263P.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
```

```
PI Jefferson RA, Kilian A, Keese PK;
XX
XX MPI: 1999-229241/19.
DR N-PSDB; AAM93826.
XX
PT New isolated microbial beta-glucuronidase.
XX
XX Example 3; Fig 13A-D; 76pp; English.
XX
CC This invention describes a novel secreted form of Bacillus sp. beta-
CC glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a
CC reporter/effector molecule and as a diagnostic tool. The products of the
CC invention can be used as markers for transgene constructions, e.g. in
CC plants or insects. They can also be used for the cleavage and
CC detoxification of glucuronides and to examine conjugation patterns of
CC glucuronides
XX
XX Sequence 602 AA;
SQ
Query Match 31.0%; Score 929; DB 2; Length 602;
Beet Local Similarity 35.6%; Pred. No. 5,1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
OY 1 MVRPQRNKKRFTLLINGVWNLV-----TSKDRPIAVPGSWNE--QYDLC 44
DB 1 MLYPIINTETRGVFDLNGVWNFKLDYKGLSEKWKESKLTDTISMAVPSYNDIGVTKR 60
OY 45 YEEGPFYKTTFFVYPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104
DB 61 NHIGVWYEREFVPAYLKDQRIVLRFGSATHKAIYYVNGELVVEHKGGLPFEAEINNS 120
OY 105 VKSENEELRVVENRKLKVGSPSKVPDSCGTHTVGFFGS-----FPANPDFFPYG 154
DB 121 LRDGNRRTVAVDNLT-----DSTLPVGLYSEHNEGKLVIRKNKNPFPPFYA 170
OY 155 GIIRPVLEFTHARLIDWDTSESSEPEKLGKVKVIESEAVGQEMTIKLG--EE 212
DB 171 GHARPVKIYTPPTFYVEDISVTVDFNGP---TGVTYTVDFQ---GKAETVKSVDDE 223
OY 213 KIRTSNRVGEFIELENARFWSLEDPYLYPLKYLEKDEYTLDI-----GIRTSWDEK 267
DB 224 GKVAVSTEGLSGNVEIPVILMEPLNTLYQIKVELVNDGLTIDVYEEFPGVRYVEVNDG 283
OY 268 RLYLNGKPVFLKGGKHEEPVLGQGTFFPLMIKDPNLTKMINANSFRTSHYPYSEEMLD 327
DB 284 KFLINNKPEYFKGFGKHEDTPINGRGFNASVMDPILKIGANSFRTAHYPYSEELMR 343
OY 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG---HLNFMATYTGEGSERVSTWEKIRTFEHHQDVLRELVS 400
OY 368 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMSMDAPDERTDV 427
DB 401 DNKHBSVVMWSIANEATEEGAYEFKPLVELTLELDPQKRPTIIVLFVMAFPE--TDK 458
OY 428 ALKPYDIVCVNRYVGYIYQGRIEBGLQLEKDIEELVYARHR---KPIFVTEFGADAIA 483
DB 455 VAEILDVIALNRYNNGYIFDGGDLBAKYLH---QEFHAMNRCCKRPMITEYGADIVA 515
OY 484 GIHYDPPEMSESYQALEVTEKTRILLKKQVYIIGTHVMAFADPKTPQVNRDILNKKGV 543
DB 516 GFHDIDIPVAFTEBEYQVEYQANHVVDFEPENFVGEQANFADPATISQGVNRVQGNKKGV 575
OY 544 TPDROPKLVANHLRLMSEV 563
DB 576 TRDRKPKLAHVFRERWTNI 595
RESULT 5
ID AAM93822 standard; protein; 602 AA.
XX
AC AAM93822;
```

```
XX
XX 25-JUN-1999 (first entry)
DT
XX
XX Bacillus sp. GUS protein.
DE
XX
XX GUS; Bogus; beta-glucuronidase; secreted; reporter molecule; marker;
XX receptor molecule; diagnostic tool; transgene construction; plant;
XX insect; cleavage; detoxification; glucuronide.
XX
XX Bacillus sp.
XX
XX WO9913085-A2.
XX
XX 18-MAR-1999.
PD
XX
XX 09-SEP-1998; 98WO-US019217.
PF
XX
XX 09-SEP-1997; 97US-0058263P.
PR
XX
XX (CAMB-) CAMBIA BIOSYSTEMS LLC.
PA
XX
XX Jefferson RA, Kilian A, Keese PK;
PI
XX MPI: 1999-229241/19.
DR
XX
XX New isolated microbial beta-glucuronidase.
XX
XX Example 1; Fig 5; 76pp; English.
XX
XX This invention describes a novel secreted form of Bacillus sp. beta-
XX glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a
XX reporter/effector molecule and as a diagnostic tool. The products of the
XX invention can be used as markers for transgene constructions, e.g. in
XX plants or insects. They can also be used for the cleavage and
XX detoxification of glucuronides and to examine conjugation patterns of
XX glucuronides
CC
CC Sequence 602 AA;
SQ
Query Match 31.0%; Score 929; DB 2; Length 602;
Beet Local Similarity 35.6%; Pred. No. 5,1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
OY 1 MVRPQRNKKRFTLLINGVWNLV-----TSKDRPIAVPGSWNE--QYDLC 44
DB 1 MLYPIINTETRGVFDLNGVWNFKLDYKGLSEKWKESKLTDTISMAVPSYNDIGVTKR 60
OY 45 YEEGPFYKTTFFVYPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104
DB 61 NHIGVWYEREFVPAYLKDQRIVLRFGSATHKAIYYVNGELVVEHKGGLPFEAEINNS 120
OY 105 VKSENEELRVVENRKLKVGSPSKVPDSCGTHTVGFFGS-----FPANPDFFPYG 154
DB 121 LRDGNRRTVAVDNLT-----DSTLPVGLYSEHNEGKLVIRKNKNPFPPFYA 170
OY 155 GIIRPVLEFTHARLIDWDTSESSEPEKLGKVKVIESEAVGQEMTIKLG--EE 212
DB 171 GHARPVKIYTPPTFYVEDISVTVDFNGP---TGVTYTVDFQ---GKAETVKSVDDE 223
OY 213 KIRTSNRVGEFIELENARFWSLEDPYLYPLKYLEKDEYTLDI-----GIRTSWDEK 267
DB 224 GKVAVSTEGLSGNVEIPVILMEPLNTLYQIKVELVNDGLTIDVYEEFPGVRYVEVNDG 283
OY 268 RLYLNGKPVFLKGGKHEEPVLGQGTFFPLMIKDPNLTKMINANSFRTSHYPYSEEMLD 327
DB 284 KFLINNKPEYFKGFGKHEDTPINGRGFNASVMDPILKIGANSFRTAHYPYSEELMR 343
OY 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG---HLNFMATYTGEGSERVSTWEKIRTFEHHQDVLRELVS 400
OY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMSMDAPDERTDV 427
```



```
Db      401 DKNHPSVWMSIANEATEBEGAYEYFKPLVELTKELDPQKRPVTVLFWMATPE--TDK 458
Qy      428 ALKYPIVCNRYGTYITVGRIEGLQALEKIDIELYARHR---KPIFTEFGADATA 483
      459 VAEILDIVIALNRNGMYFPDGDLEAAKVHLR---QEFHAWKRCPEKPIWITEYGADTVA 515
Db      484 GIHYDPQWFSSEYQALEVEKTRILLKKDYIIIGTHVMAFADKTPONRRPILNKGVF 543
      516 GFHDIDPVMFTESEYQVEYYQANHVVDEFENFVGEQANMFADPATSGVMRVQGNKKGVF 575
Qy      544 TRDRQPKLVAAHYLRRLMSEV 563
      576 TRDRKPKLAHVFRERMTNI 595
Db

RESULT 6
AAW93826 standard; protein; 602 AA.
ID      AAW93826;
XX      AAW93826;
AC      AAW93826;
XX      25-JUN-1999 (first entry)
DT
XX      Bacillus sp. GUS protein fragment.
XX      GUS; beta-glucuronidase; secreted; reporter molecule; marker;
XX      receptor molecule; diagnostic tool; transgene construction; plant;
XX      insect; cleavage; detoxification; glucuronide.
XX      Bacillus sp.
XX      WO913085-A2.
XX      18-MAR-1999.
PD      09-SEP-1998; 98WO-US019217.
XX      PF
XX      09-SEP-1997; 97US-0058263P.
PR      (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX      PA
XX      Jefferson RA, Killian A, Keese PK;
XX      PI
XX      WPI; 1999-229241/19.
XX      DR
XX      PT
XX      New isolated microbial beta-glucuronidase.
XX      PS
XX      Example 3; Fig 13D; 76pp; English.
XX      CC
XX      This invention describes a novel secreted form of Bacillus sp. beta-
XX      glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a
XX      reporter/effector molecule and as a diagnostic tool. The products of the
XX      CC invention can be used as markers for transgene constructions, e.g. in
XX      CC plants or insects. They can also be used for the cleavage and
XX      CC detoxification of glucuronides and to examine conjugation patterns of
XX      CC glucuronides
XX      SQ
XX      Sequence 602 AA;

Query Match      31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 5,1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
```

```
Db      121 LADGMNRVTVAAVDNLL-----DSTLTPVGLYSEHREGLGKVINRKNPDPFNVA 170
Qy      155 GIIRPVLIPTDHARILDIWDTSESEPEKLGKVKXIEVSEAVQGMTIKLG--EEB 212
      171 GLHRPVKVIYTFPTVEDISVYTFDENGFP--TGTVYTVDFQ---GKAETVKVSVVDBE 223
Db      213 KXIRTSNRVVEGEFPILENARFWSLEDPYLYPLKVELEKDEYTLDT-----GIRTSWDEK 267
      224 GKRVASTEGLSNVELPNVILWEPLNTLYLQIKVELVNDGLTIDVVEEPPFGVRYVEVNDG 283
Qy      268 RLYLNGKPVFLKQFGKEEFPVLQGTFFVPLMIKDNLKWINANSFRTSHYPYSEWLD 327
      284 KPLINNKPFYFKQFGHEDTPIINGKGFNANSMVDNIIKIGANSFRTAHPIYSEELMR 343
Db      328 LADRLGIIVDAAPHVGITRYAHN-----DETQXIA-----EDNIRMRIDR 368
      344 LADREGLVVIDETPAVG--HINFMATTGLGEGSERVSTWEKIRTFEHHQVLRRLVGR 400
Qy      369 HKNHPSVIMWSVANEPSNHPDAEGFKALYETANEMD-RTRPVVWVSMMDAPDEKTRV 427
      401 DKNHPSVWMSIANEATEBEGAYEYFKPLVELTKELDPQKRPVTVLFWMATPE--TDK 458
Qy      428 ALKYPIVCNRYGTYITVGRIEGLQALEKIDIELYARHR---KPIFTEFGADATA 483
      459 VAEILDIVIALNRNGMYFPDGDLEAAKVHLR---QEFHAWKRCPEKPIWITEYGADTVA 515
Db      484 GIHYDPQWFSSEYQALEVEKTRILLKKDYIIIGTHVMAFADKTPONRRPILNKGVF 543
      516 GFHDIDPVMFTESEYQVEYYQANHVVDEFENFVGEQANMFADPATSGVMRVQGNKKGVF 575
Qy      544 TRDRQPKLVAAHYLRRLMSEV 563
      576 TRDRKPKLAHVFRERMTNI 595
Db

RESULT 7
AAB28402 standard; protein; 602 AA.
ID      AAB28402;
XX      AAB28402;
AC      AAB28402;
XX      26-JAN-2001 (first entry)
DT
XX      DT
XX      Staphylococcus beta-glucoronidase.
XX      DE
XX      Microbial; beta-glucoronidase; GUS; Enterobacter; salmonella;
XX      KM Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
XX      KM transgenic insect; marker; glucuronide detoxification.
XX      OS
XX      Staphylococcus sp.
XX      PN
XX      WO200055333-A1.
XX      PD
XX      21-SEP-2000.
XX      PF
XX      16-MAR-2000; 2000WO-US007107.
XX      PR
XX      17-MAR-1999; 99US-00270957.
XX      PA
XX      (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX      PI
XX      Jefferson RA, Mayer JE;
XX      WPI; 2000-647075/62.
XX      DR
XX      N-PSDB; AAA07930.
XX      PT
XX      Novel microbial beta-glucuronidase genes and gene products used as
XX      reporter/effector molecule, as diagnostic tool, in positive selection, to
XX      target molecules to specific cells and to detect and track linked genes.
XX      PS
XX      Example 3; Fig 3; 116pp; English.
XX      CC
XX      The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
```

CC genes were obtained from six different genera: *Enterobacter*/*Salmonella*,
CC *Pseudomonas*, *Salmonella*, *Staphylococcus* and *Thermocoga*. Microbial GUS can
CC be used as a reporter/effector molecule for transgenic constructions and
CC in in vitro diagnostic applications. It may also be used to generate
CC sentinel plants that serve as bioindicators of environmental status. It
CC may be used to generate transgenic insects for tracking insect
CC populations or to facilitate the development of a bioassay for compounds
CC that affect molecules critical for insect development (e.g. juvenile
CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
CC destined for release into the environment. In animal systems, secreted
CC GUS may be used to achieve extracellular detoxification of glucuronides
CC (e.g. toxin glucuronide) and to examine conjugation patterns of
CC glucuronides. Microbial GUS may also be used in traditional medical
CC diagnostic assays, for drug testing, pharmacokinetic studies,
CC bioavailability studies, diagnosis of diseases and syndromes, following
CC progression of disease or its response to therapy. Microbial GUS has
CC increased thermal stability, high turnover number and enzymatic activity.
CC It is highly specific for the substrate and water soluble, and the
CC substrates are stable
XX

SQ Sequence 602 AA;

Query Match 31.0%; Score 929; DB 3; Length 602;

Best Local Similarity 35.6%; Pred. No. 5.1e-67;

Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```
QY 1 MVRPQNRKRFILINGVNNLEV-----TSKDRPIAVPGSANE--OYDLC 44
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MLYPINTETRGVFDINGVNNFKLDYKGLEEKWYESKLTDTISMAVPSYNDIGVTKR 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 45 YEEGPFYKTFYVPKLSQKHRLYFAAVNTDCEVLNGKVGNNHLEYLPFVEDVTGK 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 NHIGVWYERETVPAYLKQRIIVLRFGSATHKAIVYNGELVHHKGFLEFSAEINNS 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSCGTHTVGFGS-----FPANPDFPYG 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 LRDGMNRVTVAVDNLT-----DDSTLPVGLYSEHREGLGKVIKKNKPNDFFYNA 170
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 GIIRPVLEFTHARILDIWDTSSEPEPKLGKVKIYVEBAVGQEMTKLG--EE 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 171 GHRPVKYITTFYVEDISVTFDENG--TGVTYTVDFQ---GKAETVKSVNDEE 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 213 KIRTSNREVEGEFLENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 224 GKVAASTEGLSGNVEIPVILMEPLNTLYIYQIKVELVNDGLTIIDYEEFGRIVYEVNDG 283
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 268 RLYNGKRPVFLKGGKHEEPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEMLD 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 284 KFLINNKPFYFGKGGHEDTPIINGRGFNEASVNDMFNLIKIGANSFRTAHYPYSEELMR 343
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 LADRLGILVIDEAPVHGITRRHYN-----PETOKIA-----EDNIRMTDR 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 344 LADRLGILVIDEAPVHG--HUNFMATTGLGEGSERVSTWEKIRTFEHQVLELVS 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 369 HKNHPSVIMWSVANEESNHPDAEGFFKALYETANEMD--RTRPVVMSVMDAPDERTDV 427
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 401 DKNHSSVVMWSTANEAATEEBGAYEFKPLVELTLEDPQKRPVITLVFWATPE--TDK 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 428 ALKVPDIYCVNRYGYIYQRIEGLQALEKDI BELVARNR---KPIVTFEGADALA 483
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 459 VAEILDIIVIALNRYNGVYFDGDLLEAKVHLR---QEFHAMNKRCSGKPIIMITEYGADTVA 515
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 464 GIHYDRPQWSEEXQALVEKTRILLKDYIIGHVAFADPKTPONVRRIRILNHKGVF 543
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 516 GHDHIDPWFTEYQVEYIQAHHVVFDEFENFVGQAMNFADPATISQGVNRQGNKKGVF 575
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 544 TRDQPKLVAHLRLRLMSEV 563
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 576 TDRKPKLAHVRFRKMTNI 595
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 8
AAW93821

ID AAW93821 standard; protein; 618 AA.

XX AAW93821;

XX 25-JUN-1999 (first entry)

XX *Bacillus* sp. GUS protein.XX GUS; *Bogus*; beta-glucuronidase; secreted; reporter molecule; marker;

XX receptor molecule; diagnostic tool; transgene construction; plant;

XX insect; cleavage; detoxification; glucuronide.

XX *Bacillus* sp.

XX MO9913085-A2.

XX 18-MAR-1999.

XX 09-SEP-1998; 98WO-US019217.

XX 09-SEP-1997; 97US-0058263P.

XX (CAMP-) CAMBIA BIOSYSTEMS LLC.

XX Jefferson RA, Killian A, Keese PK;

XX WPI, 1999-229241/19.

XX N-PSDB; AAW93825.

XX New isolated microbial beta-glucuronidase.

XX Claim 5; Fig 4A-C; 76pp; English.

XX This invention describes a novel secreted form of *Bacillus* sp. beta-XX glucuronidase (*Bogus*). The microbial *Bogus* polypeptide can be used as a

XX reporter/effector molecule and as a diagnostic tool. The products of the

XX invention can be used as markers for transgene constructions, e.g. in

XX plants or insects. They can also be used for the cleavage and

XX detoxification of glucuronides and to examine conjugation patterns of

XX glucuronides
SQ Sequence 618 AA;

Query Match 31.0%; Score 929; DB 2; Length 618;

Best Local Similarity 35.6%; Pred. No. 5.2e-67;

Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```
QY 1 MVRPQNRKRFILINGVNNLEV-----TSKDRPIAVPGSANE--OYDLC 44
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MLYPINTETRGVFDINGVNNFKLDYKGLEEKWYESKLTDTISMAVPSYNDIGVTKR 76
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 45 YEEGPFYKTFYVPKLSQKHRLYFAAVNTDCEVLNGKVGNNHLEYLPFVEDVTGK 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 NHIGVWYERETVPAYLKQRIIVLRFGSATHKAIVYNGELVHHKGFLEFSAEINNS 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSCGTHTVGFGS-----FPANPDFPYG 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 137 LRDGMNRVTVAVDNLT-----DDSTLPVGLYSEHREGLGKVIKKNKPNDFFYNA 186
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 GIIRPVLEFTHARILDIWDTSSEPEPKLGKVKIYVEBAVGQEMTKLG--EE 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 187 GHRPVKYITTFYVEDISVTFDENG--TGVTYTVDFQ---GKAETVKSVNDEE 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 213 KIRTSNREVEGEFLENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 240 GKVAASTEGLSGNVEIPVILMEPLNTLYIYQIKVELVNDGLTIIDYEEFGRIVYEVNDG 299
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 268 RLYNGKRPVFLKGGKHEEPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEMLD 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 300 KFLINNKPFYFGKGGHEDTPIINGRGFNEASVNDMFNLIKIGANSFRTAHYPYSEELMR 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 LADRLGILVIDEAPVHGITRRHYN-----PETOKIA-----EDNIRMTDR 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

DB 360 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHNDVLRBLVSR 416
QY 369 HGNHPSVIMSVANEPESNHPDAEGFFKLLYETANMD-RTRPVVMSMDAPDETRTV 427
DB 417 DKNHPSVIMSVANEPESNHPDAEGFFKLLYETANMD-RTRPVVMSMDAPDETRTV 474
QY 428 ALKYPFIVCVNRYGYIYQGRIEBGLQLEKDIIEELVARRH---KPIFVPEFGADTA 483
DB 475 VAEIDIVIALNRKNGYFPGGDLLEAKVHLR---QEFHAMKRCPCGKPIMTIEYGADTVA 531
QY 484 GIHYDPQMFSEBYQAELEVEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF 543
DB 532 GFHDIDPVMTIEEYQVEYVQANHVPDEFENFVGEQAMNFPADPATSGVMRVQGNKKGVF 591
QY 544 TDRDQPKLVAVHLRLMSEV 563
DB 592 TDRKPKLAHVFRERWTNI 611

RESULT 9
AAB28408
ID AAB28408 standard; protein; 615 AA.
AC AAB28408;
XX
XX
DT 26-JAN-2001 (first entry)
DE Codon-optimised Staphylococcus GUS.
XX
XX Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KM Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KM transgenic insect; marker; glucuronide detoxification.
XX
XX Staphylococcus sp.
OS
XX
XX WO200055333-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 16-MAR-2000; 2000MO-US007107.
PF
XX
XX 17-MAR-1999; 99US-00270957.
PR
XX
XX (CAMB-) CAMBIA BIOSYSTEMS LLC.
PA
XX
XX Jefferson RA, Mayer JE;
PI
XX
XX WPI; 2000-647075/62.
DR
XX
XX N-PSDB; AAA07938.
DR

Novel microbial beta-glucuronidase genes and gene products used as reporter/effector molecule, as diagnostic tool, in positive selection, to target molecules to specific cells and to detect and track linked genes.

Claim 3; Fig 5B; 116pp; English.

The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS genes were obtained from six different genera: Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can be used as a reporter/effector molecule for transgenic constructions and in in vitro diagnostic applications. It may also be used to generate sentinel plants that serve as bioindicators of environmental status. It may be used to generate transgenic insects for tracking insect populations or to facilitate the development of a bioassay for compounds that affect molecules critical for insect development (e.g. juvenile hormone). Secreted GUS may also serve as a marker for beneficial fungi designed for release into the environment. In animal systems, secreted GUS may be used to achieve extracellular detoxification of glucuronides (e.g. toxin glucuronide) and to examine conjugation patterns of glucuronides. Microbial GUS may also be used in traditional medical diagnostic assays, for drug testing, pharmacokinetic studies, bioavailability studies, diagnosis of diseases and syndromes, following progression of disease or its response to therapy. Microbial GUS has

CC increased thermal stability, high turnover number and enzymatic activity.
CC It is highly specific for the substrate and water soluble, and the
CC substrates are stable
XX
XX Sequence 615 AA;
SQ

Query Match 30.8%; Score 925; DB 3; Length 615;
Best Local Similarity 35.7%; Pred. No. 1,1e-66;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

4 PQRNKKRPFILLNGVNNLEV-----TSKRPIAVPGSWE--QYDLCYBE 47
DB 9 PINTETRGVDFNDGWNPFRLDVGKLEBKWYESKLTDTISNAVPSYNDIGTKEIRNHI 68
QY 48 GPFTYKTTFFYDPKLSQKHRLYPAAVNTDCVPLNGEKVGENHIEYLPEVDVTGKYS 107
DB 69 GYWMYERBEFTVPAVLDQRIYVRFSSATHKALVYVNGELVHKKGFPLFEAEINNSLD 128
QY 108 GENELRVVENVRLKYGFPFSKVPDSCGTHVGFSG-----FPANPDFPYGGII 157
DB 129 GNNRVTVAVDNTL-----DSTLPVGLYSERHBEGLAKYIRNKNPFDFVYAGLI 178
QY 158 RPLIEFTDHAIRLIDVWDTSESEBEKLGKVKVIEVSEAVQEMTIKLG--BEEKKI 215
DB 179 RPKLYTTPFTYVEIDISVTDPNGR---TGVTYTVDPQ---GKAETVKVSVNDEBGKV 231
QY 216 RTSNRVGEFTLENARFWSLEDPYLYPLKVELEKDEYTLDI----GIRTSMDKRLY 270
DB 232 VASTEGLSGNVPIPVVILMEPLNTLYQIKVELVNDGLTIIDYEBPFGVTVAVNDGKPL 291
QY 271 LNKGVPLKPGKHEEPVLQGTTPYPLMIDPNLKKTNNSRPTSHYPSEELDLAD 330
DB 292 INKKPFYEGFGEKHEDTPIINGRFGNEASVNDENILKMGANSFTAHYPSEELRLAD 351
QY 331 RLGIIVIDEAPHVIGTRVHYN-----PETOKLA-----EDNIRRMIDRHKN 371
DB 352 REGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHNDVLRBLVSRDN 408
QY 372 HPSVIMSVANEPESNHPDAEGFFALYETANEMD-RTRPVVMSMDAPDETRTVALK 430
DB 409 HPSVIMSVIANEAATEEGAYEFKPLVELTKELDPOKRPVITVLFVMTPE--TDKVAE 466
QY 431 YFDIVCVNRYGYIYQGRIEBGLQLEKDIIEELVARRH---KPIFTEPADIAIGH 486
DB 467 LIDVIALNRKNGYWDGDLLEAKVHLR---QEFHAMKRCPCGKPIMTIEYGADTVAGH 523
QY 487 YDPQMFSEBYQAELEVEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVTRD 546
DB 524 DIDPVMTIEEYQVEYVQANHVPDEFENFVGEQAMNFPADPATSGVMRVQGNKKGVFTRD 583
QY 547 RQPKLVAVHLRLMSEV 563
DB 584 RKPKLAAHVFRERWTNI 600

RESULT 10
AAB28409
ID AAB28409 standard; protein; 602 AA.
AC AAB28409;
XX
XX
DT 26-JAN-2001 (first entry)
DE Salmonella beta-glucuronidase.
XX
XX
XX Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KM Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KM transgenic insect; marker; glucuronide detoxification.
XX
XX Salmonella sp.
OS
XX
XX WO200055333-A1.
PN

PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000MO-US007107.
 XX
 PR 17-MAR-1999; 99US-00270957.
 XX
 PA (CAMP-) CAMBIA BIOSYSTEMS LLC.
 XX
 PI Jefferson RA, Mayer JE;
 XX
 DR WPI, 2000-647075/62.
 DR N-PSDB; AAA07939.
 XX
 PT Novel microbial beta-glucuronidase genes and gene products used as
 PT reporter/effector molecule, as diagnostic tool, in positive selection, to
 PT target molecules to specific cells and to detect and track linked genes.
 XX
 PS Example 3; Fig 17; 116pp; English.
 XX
 CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
 CC genes were obtained from six different genera: Enterobacter/salmonella,
 CC Pseudomonas, salmonella, staphylococcus and thermotoga. Microbial GUS can
 CC be used as a reporter/effector molecule for transgenic constructions and
 CC in vitro diagnostic applications. It may also be used to generate
 CC sentinel plants that serve as bioindicators of environmental status. It
 CC may be used to generate transgenic insects for tracking insect
 CC populations or to facilitate the development of a bioassay for compounds
 CC that affect molecules critical for insect development (e.g. juvenile
 CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
 CC destined for release into the environment. In animal systems, secreted
 CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies,
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable
 XX
 SQ Sequence 602 AA;
 Query Match 30.7%; Score 920; DB 3; Length 602;
 Best Local Similarity 35.3%; Pred. No. 2.8e-66;
 Matches 213; Conservative 110; Mismatches 223; Indels 58; Gaps 11;
 QY 1 MRPQRNKKRFFILLNGVNLVETSKD-----RPIAVGSWNEQY--QDL 43
 DB 1 MRSVETATREIKLDGLMSPCMDSCECNAQOMWRQPLRPOSRAIAVPGSYNDQFAAAEI 60
 QY 44 CYEEGPFYKTFYVVKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVYG 103
 DB 61 RNYVNVWYQREIRIPKGDROIRIVLPFAVTHYKGVWVNDQFLMEHOGCTYFEEADISH 120
 QY 104 KYKSGEN-ELRVVENRLKVGSPSKVPSGTHVGFSGSPRPANF-DFPYGGIIRPVL 161
 DB 121 LISAGSVYITVCVNNELMWQTRP-----PGVVTQGVNKGQQAAYFHDFNFVAAGIHRSM 175
 QY 162 IEFTHARLIDIVDISESEPEKKGAKVKIYVSEAVGQEMTILGSEEEKIKITSNF 221
 DB 176 LVTTPKTFEEDITVTVQVAD---DLAQATVAMQVRANG---EVRVELRDAEQOLVAASGGG 229
 QY 222 VEGEFLLENARFMSLEDPLYLPLKV---ELEKDEVTLDIGRTISWDEKRLYLNGKPYF 277
 DB 230 EKGELLLEBGRIMQPEGILYELRYAQAQODEDEYPLAVGIRSVKGEQGLINHKRPY 289
 QY 278 LKFGKGEFFVLQGTFFYPLMIKQFNLLKWINANSFRITSHYVSEBMLDLRLGILVI 337
 DB 290 FFGFGHEDADLRKSGFNVLMVHDALMDWIGANSYRTSHRYAEMLDMADEHGIVLI 349
 QY 338 DEAPHVGT-----TRYHNPETQKLAEDNIRKIMIDRHKNHPSVIMS 379
 DB 350 DETAAVGFNLISGISFVGEKPEKELYDEAVNDETQRAHLQAIKELIARDKNHPSVWMS 409

QY 380 VANEPESHNPDAEGFKALEYETANEMDRTRPVVWSKM--DAPDERTDVALKPYDIYCV 437
 DB 410 IANEPDTRPNQARREYFAPLAQATRELDPTRTITCVNVWFCDHSESTITDL----FDVYCL 465
 QY 438 NRYYGWYIYQRIEGLQALEKDIIEELYARRKPIFYTFEGADALAGIHYDPQWFSSEY 497
 DB 466 NRYYGWYVQSGDLEKAEKYLEKELLAMQEKHRRPIITREYGVDTLAGHSWYNDWSESEY 525
 QY 498 QAEIVKTIIRLLKKDYIIGTHWAFADPKTPQNVRRPILNKKGVFTDRQPKLVAYLVR 557
 DB 526 QCAMLMDYHVRVDRVASAVVGEQVWVFADPATSOGIMRVGNKKGIPTDRRKRKAAPLLQ 585
 QY 558 RLMS 561
 DB 586 KRWI 589
 RESULT 11
 AAM48998
 ID AAM48998 standard; protein; 598 AA.
 XX
 AC AAM48998;
 XX
 DT 10-MAY-2002 (first entry)
 DT
 DE L gasseri beta-glucuronidase.
 XX
 KW Beta-glucuronidase; GUS; low pH; Lactobacillus.
 XX
 OS Lactobacillus gasseri.
 XX
 PN W0200190305-A2.
 PD 29-NOV-2001.
 PF 22-MAY-2001; 2001MO-US016667.
 PR 23-MAY-2000; 2000US-0206372P.
 PA (UNIC-) UNIV NORTH CAROLINA STATE.
 XX
 PI Russell WM, Klaenhammer TR;
 XX
 DR WPI; 2002-062529/08.
 DR N-PSDB; AAL44836.
 PT Novel isolated nucleic acid of Lactobacillus gasseri encoding beta
 PT glucuronidase having activity at acidic pH which is useful as reporter
 PT protein for highly aciduric organisms, or as marker of transformed cells.
 PS Claim 19; Page 49-51; 53pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC Lactobacillus gasseri beta-glucuronidase (GUS). The protein has maximum
 CC activity at acidic pHs, and thus can be used to detect low pH
 CC environments. The present sequence is the protein of the invention
 XX
 SQ Sequence 598 AA;
 Query Match 30.6%; Score 919.5; DB 5; Length 598;
 Best Local Similarity 34.1%; Pred. No. 3e-66;
 Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;
 QY 4 PQRNKKRFFILLNGVNLVET-----SKRP---IAYGSWNE--QYDLCYEE 47
 DB 7 PIONKYRENTLMNGWQETDPNSVGLDEGNKKELPDEBNRPVGTFAELTTKRDRKYYT 66
 QY 48 GPFYKTFYVVKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVYGKYS 107
 DB 67 GDFWYQKDFIFSPFKKELIYRGSVYHRAKVPFINGEVQDHGCGFLPFQVKLSNYNY 126
 QY 108 GENELRVVENRLKVGSP--SKVPDSGTHVGFSGSPRPANFDFPYGGIIRPVLE 163

Qy 414 VSMMDAPDERTRDVALKYPFDIVCVNRYGMYIYQGRIEBGLQALEKDI ELYA--RHRK 470
Db 473 AIAV----SNTQDKAGRSJDIISFNRYNMYSNAGRLD--MITQWVIDEALANRKRYNK 525
Qy 471 PIFVTEFGADAIAGIHYPDPQMFSEBYQALVEKTI RL--LLKDYIIGTHVMAFADPK 527
Db 526 PIMSEYGADTLEGLHMQPAYVWSEEFQTEVFPSRHPKAFDELRLKKGWFIGEFVNNFADPK 585
Qy 528 TPQNVRRPILNKHGVPTRDROPKLVAVHLRR 558
Db 586 TAQSYTRVGNGKKGVFTRARQPPAAAHLLRK 616

Search completed: September 1, 2005, 19:29:37
Job time : 172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:22:34 ; Search time 42 Seconds

(without alignment)
1289.763 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001

Sequence: 1 MVRPQNRKRFLLINGVWN.....TRDRQPKLVAVHRLRMSRV 563

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	2 A72300	beta-glucuronidase
2	1011	33.7	570	2 C90485	beta-glucuronidase
3	899.5	30.0	648	2 A32576	beta-glucuronidase
4	894.5	29.8	648	2 A25047	beta-glucuronidase
5	879	29.3	603	1 GBEGC	beta-glucuronidase
6	865.5	28.8	651	2 A26581	beta-glucuronidase
7	691	23.0	368	2 A85768	partial beta-D-glu
8	691	23.0	370	2 D90919	beta-D-glucuronida
9	487	16.2	755	2 D95842	probable beta-gala
10	482	16.1	716	2 JU0275	beta-galactosidase
11	481.5	15.0	996	2 D86872	beta-galactosidase
12	452.5	15.1	1087	2 F72283	beta-galactosidase
13	414	13.8	1024	1 GBEGC	beta-galactosidase
14	405.5	13.5	1024	2 E90678	beta-galactosidase
15	405.5	13.5	1024	2 A85529	beta-D-galactosida
16	381.5	12.7	1014	2 C83990	beta-galactosidase
17	371.5	12.4	1307	2 T35944	probable beta-gala
18	364	12.1	897	2 A39405	beta-galactosidase
19	360.5	12.0	1060	2 A10201	beta-galactosidase
20	352.5	11.7	1025	1 JCI266	beta-galactosidase
21	352.5	11.7	1042	1 GBEGC	beta-galactosidase
22	352.5	11.7	1042	2 E85968	evolved beta-D-gal
23	352.5	11.7	1042	2 P91123	evolved beta-D-gal
24	348.5	11.6	1034	2 T30574	beta-galactosidase
25	348	11.6	1026	2 A49750	beta-galactosidase
26	339.5	11.3	1034	2 A24925	beta-galactosidase
27	328	10.9	1075	2 T47603	beta Galactosidase
28	327.5	10.9	1034	2 T30551	beta-galactosidase
29	322.5	10.7	626	2 A42891	beta-galactosidase

30	321	10.7	2228	2 E97942	beta-galactosidase
31	321	10.7	2233	2 B95075	beta-galactosidase
32	312	10.4	1015	2 I39697	beta-galactosidase
33	310	10.3	1005	2 T31333	beta-galactosidase
34	300.5	10.0	1007	2 A30093	beta-galactosidase
35	256	8.5	785	2 H72228	hypothetical prote
36	194	6.5	237	2 E90919	interrupted beta-D
37	194	6.5	237	2 B85768	partial beta-D-glu
38	191.5	6.4	827	2 AB2764	beta-mannosidase p
39	191.5	6.4	832	2 A97545	hypothetical prote
40	181	6.0	820	2 T37230	probable beta-mann
41	179.5	6.0	900	2 T19689	hypothetical prote
42	179	6.0	879	2 A55881	beta-mannosidase (
43	175	5.8	891	2 AB2755	beta-mannosidase p
44	139.5	4.6	1091	2 F83928	hypothetical prote
45	132	4.4	818	2 A97668	mannosidase AGR_C_

ALIGNMENTS

RESULT 1					
A72300 beta-glucuronidase - Thermotoga maritima (strain MSB)					
C:Species: Thermotoga maritima					
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004					
C:Accession: A72300					
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.					
Nature 399, 323-329, 1999					
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq					
A:Reference number: A72300					
A:Accession: A72300					
A>Status: preliminary					
A:Molecule type: DNA					
A:Residues: 1-563 <ARN>					
A:Cross-references: UNIPROT:Q9X0F2; GB:AE001766; GB:AE000512; NID:g4981600; PIDN:AAD3614					
A:Experimental source: strain MSB					
C:Genetics:					
A:Gene: TM1062					
C:Superfamily: beta-glucuronidase					
Query Match					
Best Local Similarity 99.8%; Pred. No. 4.2e-197;					
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1		1		60
DB	1	MVRPQNRKRFLLINGVWNLEVTSKDRPIAVPGSMNOYODLCYEEGPFYKTFYVPK	60		
QY	61	XLSQGHIRLYFAAVNTDCEVFLNGEKVGNHIEVLPFEVDVYTKVSGENELRVVENVRL	120		
DB	61	ELSQGHIRLYFAAVNTDCEVFLNGEKVGNHIEVLPFEVDVYTKVSGENELRVVENVRL	120		
QY	121	KVGFPSKVPDGGTHTVGFPGSFPFANPDPFPGYGIIRPVLIETDARILDIWDTSES	180		
DB	121	KVGFPSKVPDGGTHTVGFPGSFPFANPDPFPGYGIIRPVLIETDARILDIWDTSES	180		
QY	181	BEKKLGKVKYKIVSEEAIVGEMTKLGBEEKKIRTNRFVEGFILLENAPFMSLEDPY	240		
DB	181	BEKKLGKVKYKIVSEEAIVGEMTKLGBEEKKIRTNRFVEGFILLENAPFMSLEDPY	240		
QY	241	LYPLKVELEKEVTLDIGIRITSWDEKRLYNGKVPFLKGFQKHEFPVLGGSTFYPLMT	300		
DB	241	LYPLKVELEKEVTLDIGIRITSWDEKRLYNGKVPFLKGFQKHEFPVLGGSTFYPLMT	300		
QY	301	KDFNLKWINANSFRTSHYPISEEWLADRLGILVIDEAPHVGTTRHYNPETOKIAD	360		
DB	301	KDFNLKWINANSFRTSHYPISEEWLADRLGILVIDEAPHVGTTRHYNPETOKIAD	360		
QY	361	NIRRMIDHKKHNPSTVIMNSVANEPSNHPDGEFFKALYETANEMDRPRPVVMSMDAP	420		
DB	361	NIRRMIDHKKHNPSTVIMNSVANEPSNHPDGEFFKALYETANEMDRPRPVVMSMDAP	420		

Db 361 NIRMIDRKHNSVIMVANEPSNHPDAGEFKALYETANEMDRTRPVVWMSMDAP 420
Qy 421 DERTDVALKYPDIYVNVNYYGMYIYQGIIEGLQALEDEIDELVARNHKEPIFTEFGAD 480
Db 421 DERTDVALKYPDIYVNVNYYGMYIYQGIIEGLQALEDEIDELVARNHKEPIFTEFGAD 480
Qy 481 AIAGIHYDDPQMFSESEYQALVEKTRILLLKKDYIIGTHVMAFADFKTEQVNRRLNNK 540
Db 481 AIAGIHYDDPQMFSESEYQALVEKTRILLLKKDYIIGTHVMAFADFKTEQVNRRLNNK 540
Qy 541 GVFTDRQPKLVAAHVLRLMSEV 563
Db 541 GVFTDRQPKLVAAHVLRLMSEV 563
RESULT 2
beta-glucuronidase (gnab) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: C90485
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweyer, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
attreit, R.A.; Ragan, M.A.; Sensesen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: C90485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-570 <KUR>
A:Cross-references: UNIPROT:Q97U11; GB:AE006641; NID:g13816434; PIDN:AAK43138.1; GSPDB:G
C:Genetics:
A:Gene: gusB
C:Superfamily: beta-glucuronidase
Query Match 33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 2.7e-61;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;
Qy 15 LMGVNLVETSKDRP-----IAPGSMNEQYQDLCYEEGFTYKTTFYVPK 60
Db 11 LOGFMFKKIDNTEGSENGMYKGLSESDIIVPASWNEQNPWQDFSGIAWYQKDLFVSN 70
Qy 61 XLSQKIRIRYPAAVNNDCEVPLNGEKVGNHIEYLFEVDVYKSKSGNELRVVVENRL 120
Db 71 DNGNRKAMVFBGAGYITLKLWINGEGTHESFTQFKPIKLV---NEFNKTV---V 123
Qy 121 KYGFPSPKVPDSGTHTVGFFGSPFPAN-----FDFFPYGGIIRPVLEFTDHARILDIW 174
Db 124 KIDNTPSPY-----NLPPARDLNNAFDFPNVGGIHRPVYIEFVDECHVEDIT 171
Qy 175 VDTSESEPKKLGKVKVKEVSEAVGQEMTIKLGEEKKI---RTSNRVEGEFIEEN 230
Db 172 VVT-----KSYGHLKVEI-LSECNQRFSLRFVLVDEKGRVILNNESSNEVEEKO--VNN 222
Qy 231 AAFWLSDEDFYLPKLELE-----KDEYTLDIGIRTSIDMKRYLNGKVPVLKGGKKE 285
Db 223 VTPWSDNDFYLLTYLWEMVVGKLVSYERIGFRVVEVXKDGITVILNGRPITLKGGRIE 282
Qy 286 EFPVLQGTFFYPLMKIDFNILKMINANSFRTSHYPSSEMLDLADRLGILVDEAP--HV 343
Db 283 DFFILGKFTYGAVLVADFLLMRKIGANSFRTSHYPSNSHLDLADMGFLVLEFPPLCS 342
Qy 344 GTRHYNDPTOKI-----AEDNIRRMIDRKHNSVIMVWVANEPSNHPDAGEF 394
Db 343 NISRWVSEIEIAKMFEDVKYFEKVRDTIKEMIRQHNKRPVIMVSWNEPDDIEVAEF 402
Qy 395 FKALYETANEMDRTRPVVWMSMDAPDERTDVALKYPDIYVNVNYYGMYIYQGIIEGL 454
Db 403 IREVELFKSLDSRSRVTPTAS-----HRSVRDLALEYVIVISLNTYHGYTMEGDISGV 457
Qy 455 QALEKDIIELVARN-RKPIFVTEFGADALAGIHYDDPQMFSESEYQALVEKTRILLLKKD 513

Db 458 KVALIELEIHKKPEKRIITTEFGADALVGLHSDPPQMFSESEYQSEMIKRYIETALREKD 517
Qy 514 YIIGTHVMAFADFKTPQVNRRLNNKGVFTDRQPKLVAAHVLRLM 560
Db 518 YIVGFHINWFADFRTPQNPSTRILNRKGIFFTRDRQPKLVAAHVLRLM 564
RESULT 3
A32576
beta-glucuronidase (EC 3.2.1.31) allele B precursor - mouse
N:Alternate names: beta-D-glucuronoside glucuronosylhydrolase
C:Species: Mus musculus (house mouse)
C>Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004
C:Accession: A32576; T49692; A28954; A29977; A35738
R:Marzynaik, C.J.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.B.; Lund, S.D.; Rinchik, E.
Mol. Cell. Biol. 9, 4074-4078, 1989
A>Title: DNA determinants of structural and regulatory variation within the murine beta-
A:Reference number: A32576; MUID:89384641; PMID:2779578
A:Accession: A32576
A:Molecule type: mRNA
A:Residues: 1-648 <MA>
A:Cross-references: UNIPROT:P12265; GB:M28540; GB:M2816; NID:g193718; PIDN:AAA63307.1;
A:Experimental source: allele B
A:Accession: B32576
A:Molecule type: mRNA
A:Residues: 1-86, 'I', 88-648 <MA2>
A:Cross-references: GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:g193721; GB:M2816
A:Experimental source: allele H
R:Funkenstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.
Mol. Cell. Biol. 8, 1160-1168, 1988
A>Title: Genomic organization and sequence of the Gus-B-a allele of the murine beta-gluc
A:Reference number: 149692; MUID:88216590; PMID:2835664
A:Accession: 149692
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-232, 'T', 234-264, 'D', 266-319, 'V', 321-427, 'K', 429-615, 'L', 617-648 <RES>
A:Cross-references: GB:M9279; NID:g193524; PIDN:AAA37697.1; PID:g309257
A:Experimental source: allele A
R:D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.
Biochemistry 27, 7131-7140, 1988
A>Title: Complete sequence and organization of the murine beta-glucuronidase gene.
A:Reference number: A28954; MUID:89062453; PMID:3196706
A:Accession: A28954
A:Molecule type: DNA
A:Residues: 1-264, 'D', 266-319, 'V', 321-648 <DM>
A:Cross-references: GB:J02836; GB:U03035; GB:M20204; NID:g193716; PIDN:AAA98623.1; PID:g
R:Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.
Genomics 2, 215-219, 1988
A>Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its dedu
A:Reference number: A29977; MUID:88284700; PMID:3397060
A:Accession: A29977
A:Molecule type: mRNA
A:Residues: 1-264, 'D', 266-319, 'V', 321-648 <GAL>
A:Cross-references: GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:g309256
R:Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.
J. Biol. Chem. 265, 14732-14735, 1990
A>Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in co
the serpin superfamily.
A:Reference number: A35798; MUID:90368633; PMID:2394691
A:Accession: A35798
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 593-648 <LIA>
A>Note: the location of the propeptide cleavage site was not demonstrated directly but r
C:Comment: in some tissues, a portion of this enzyme is retained in the endoplasmic ret
syn.
C:Genetics:
A:Gene: Gus
A:Map position: 5
A:Introns: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1
C:Superfamily: beta-glucuronidase
C:Keywords: glycosidase; hydrolase; lysosome

F.1-22/Domain: signal sequence #status predicted <SIG>
 F.23-648/Product: beta-glucuronidase, BR-retained form #status predicted <EMT>
 F.23-633/Product: beta-glucuronidase, lysosomal (default) form #status predicted <LMAT>
 F.634-648/Domain: carboxyl-terminal propeptide #status predicted <CPRO>

Query Match 30.0%; Score 899.5; DB 2; Length 648;

Best Local Similarity 35.9%; Pred. No. 1.3e-53;

Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

```

QY 1 MRPQRNKKRFLIILNGVNNLEV-TSKDR-----PIAVGSGWNEQ 39
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 27 MLFPKSPERELKALDGLWHFRADLSNNRLQGFEQGWYRQPLRESGPLMDPSSPFNDI 86
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 40 YODLCEE--GPTTYKTTFFVVPKXLSQ---KHRLKFAAVNTDCEVPLNGEKVGENHLEY 94
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 87 TQEAALRDTIGWVWYEREAAILPRRWQDTDRVVLRLNSAHYAAVWVWGVHVEHEG 146
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 95 LPEFVDVTGKVKSG---ENELRVVENRLKVGGFPSKVP-----DSGHTVGFPGSF 143
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 147 LPEFADISLTVQSGPLTTCRITAINNTLT---PHTLPRTIVYKTDPSMKGYF--V 200
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 144 PRANDPFPYGGIIRPVLIPTDHAIRLIDIWDTSESEBEKLGKVK--VKIEVSEAVG 201
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 201 QDTSDFPFYAGIHRSVLYTPTTYIDITVITN---VEQDIGLVTWISVQSGSEHF-- 255
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 202 QEMTIKLGSEBEKKIRTSNRPVGEFPLENARFW---SLEDP-VLYPLKV-----ELEKD 251
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 256 -LLEVQLDDEGGVVAHGTGNOGQLVPSANLWMPYLMHEHPAYWYSLVKVTTESVTD 314
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 252 EYTLDIGRTISWDEKRLYLNGKPVFLKGFGRKHEEPVLGQGFYPLMIKDNLKMINA 311
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 315 YTLPLGRTIVAVTKSPFLNGKRPFRQGVNKHEDIDIRKGFMDLLVKDFNLRLMQLA 374
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 312 NSFRTSHVPYSEEWLADLRGLIVIDEAPHVGIT-----RYHNPETOKIAEDN 361
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 375 NSFRTSHVPYSEEVQLCDRYGIVIVDECPGVGI VLPQSGNBSLRHHL-----EVMDELVR 429
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 362 IRRMIDRHKHNPVSIMVSNANPEBESNHPDAEGFKALYETANEMDRTRPVVWVS--MMA 419
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 430 VR---RDKNHRAVWVMSVANEBSLSLKPAAYFKTLIHTKALDLTRPVTVSNAXYDA 485
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 420 PDERTDVALKTFDVCVARYGYMYTYOGRIEGLQALEKDI BELYARRKPIFVTEFGA 479
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 486 -----DLGAPYVDVLCVNSYFSWHDYGHLEVIQQLNSQFENWYKTKOKPIIOSEYGA 539
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 480 DALIAGIHYPDPQMFSESEYQAELEVEKTIKRL--LKDYIIGTHVWAPADPKTPQNVRRPIL 537
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 540 DALPGIHEDPPRMFSESEYQKAVLENYHSLVDOKRKEVYVGGELIWNFADPMTQSPRLRYIG 599
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 538 NHHGVFTRDRQPKLVAVHVR-RLW 560
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 600 NKGIFTRQRPKTSAFILRERYW 623
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 4

A25047
 beta-glucuronidase (EC 3.2.1.31) precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004
 C:Accession: A25047; S00345
 R:Nishimura, Y.; Rosenfeld, M.G.; Kreibich, G.; Gubler, U.; Sabatini, D.D.; Adegnik, M.; Proc. Natl. Acad. Sci. U.S.A. 83, 7292-7296, 1986
 A:Title: Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro
 A:Reference number: A25047; MUID:87016933; PMID:3463967
 A:Accession: A25047
 A:Molecule type: mRNA
 A:Residues: 1-648 <NIS>
 A:Cross-references: UNIPROT:P06760; GB:M13962; NID:G204329; PIRN:AAA1228.1; PID:G204330
 A:Experimental source: female preputial gland
 R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S. Biochem. J. 250, 547-555, 1988
 A:Title: Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and expression
 A:Reference number: S00345; MUID:88183378; PMID:3355537

A:Accession: S00345
 A:Molecule type: mRNA
 A:Residues: 'E', '15-20', 'L', '22-486', 'L', '488-648' <PON>
 A:Cross-references: EMBL:Y00717; NID:G56270; PID:CAA68705.1; PID:G758260
 C:Superfamily: beta-glucuronidase
 C:Keywords: glycosidase; hydrolase
 F.1-22/Domain: signal sequence #status predicted <SIG>
 F.23-648/Product: beta-glucuronidase #status predicted <MAT>

Query Match 29.8%; Score 894.5; DB 2; Length 648;

Best Local Similarity 36.1%; Pred. No. 2.9e-53;

Matches 225; Conservative 98; Mismatches 210; Indels 91; Gaps 20;

```

QY 1 MRPQRNKKRFLIILNGVNN-----LEVTSKQRPVAVSGWNE- 38
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 27 MLFPKSPERELKALDGLMSFPADYSNNRLQGFEKQWYRQPLRESGPLTMDPSSPFNDI 86
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 39 -QYODLCEE--GPTTYKTTFFVVPKXLSQKHRLKFAAVNTD---CEVPLNGEKVGENHLEY 94
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 87 TQEAALRDTIGWVWYEREAAILPRRWQDTDRVVLRLNSAHYAAVWVWGVHVEHEG 146
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 95 LPEFVDVTGKVKSG---ENELRVVENRLKVGGFPSKVP-----DSGHTVGFPGSF 143
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 147 LPEFADISLTVQSGPLTTCRITAINNTLT---PYTLPRGTIVYKTDPSMKGYF--V 200
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 144 PRANDPFPYGGIIRPVLIPTDHAIRLIDIWDTSESEBEKLGKVKKIEVSEAVGQ 203
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 201 QDISDFPFYAGIHRSVLYTPTTYIDITVITN---DVDKDVGLVNWVISV-QGSDHPO 256
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 204 MTKLGSEBEKKIRTSNRPVGEFPLENARFW---SLEDP-VLYPLKV-----ELEKDEY 253
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 257 LSVRLLEDGKTVANGTGNBGLKVPRAHLMWPMYLMHEHPAYWYSLVKVTTESVTDY 316
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 254 TLDIGRTISWDEKRLYLNGKPVFLKGFGRKHEEPVLGQGFYPLMIKDNLKMINNS 313
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 317 TLPGVIRVAVTVKSPFLNGKRPFRQGVNKHEDSDIRKGFMDLLVKDFNLRLMQLANS 376
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 314 FRTSHVPYSEEWLADLRGLIVIDEAPHVGIT-----RYHNPETOKIAEDNIR 363
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 377 FRTSHVPYSEEVQLCDRYGIVIVDECPGVGIVLPSQSGNLSLRHHL-----EVMDELVR 431
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 364 RMIDRHKHNPVSIMVSNANPEBESNHPDAEGFKALYETANEMDRTRPVVWVSMDAPDR 423
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 422 ----RDKNHRAVWVMSVANEBSLSLKPAGYFKTLIAHTKALDTRPVTVS-----N 480
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 424 TR---DVALKTFDVCVARYGYMYTYOGRIEGLQALEKDI BELYARRKPIFVTEFGAD 480
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 481 TRYDADMGAPYVDVLCVNSYFSWHDYGHLEVIQQLNSQFENWYKTKOKPIIOSEYGD 540
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 481 ALIAGIHYPDPQMFSESEYQAELEVEKTIKRL--KKDYIIGTHVWAPADPKTPQNVRRPIL 537
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 541 AVSGIHEDPPRMFSESEYQATLLE-NYHILDEKREKRYVIGELIWNFADPMTQSPRLRYIG 599
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 538 NHHGVFTRDRQPKLVAVHVR-RLW 560
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 600 NKGIFTRQRPKMAAFILRERYW 623
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 5

GBRCGC
 beta-glucuronidase (EC 3.2.1.31) uidA - Escherichia coli (strain K-12)
 N:Alternate names: beta-D-glucuronoside glucuronosylhydrolase; guuA protein
 C:Species: Escherichia coli
 C:Date: 30-Jun-1988 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: C64918; I53717; A26487; S43555
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: C64918
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-603 <BLAT>
A:Cross-references: UNIPROT:P05804; GB:A0000257; GB:U00096; NID:g1787898; PIDN:AACT4689.
A:Experimental source: strain K-12, substrain MG1655
R.Schlamann, H.R.; Riiseeuv, E.; Franke-van Dijk, M.E.; Hooykaas, P.J.
Gene 138, 259-260, 1994
A:Title: Nucleotide sequence corrections of the uida open reading frame encoding beta-galactosidase
A:Reference number: 153717; MUID:94171050; PMID:8125312
A:Accession: 153717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-603 <RES>
A:Cross-references: GB:S69414; NID:G545893; PIDN:AA30197.1; PID:G545894
R.Jefferson, R.A.; Burgess, S.M.; Hirsch, D.
Proc. Natl. Acad. Sci. U.S.A. 83, 8447-8451, 1986
A:Title: Beta-glucuronidase from Escherichia coli as a gene-fusion marker.
A:Reference number: A26487; MUID:87041472; PMID:3534890
A:Accession: A26487
A:Molecule type: DNA
A:Residues: 1-419, 'VHGNI', 427-603 <JEP>
A:Cross-references: GB:M14641; NID:G868017; PIDN:AAA68923.1; PID:G868020
R.Punt, P.
submitted to the EMBL Data Library, April 1994
A:Reference number: S43555
A:Accession: S43555
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1, 'V', 3-603 <PUN>
A:Cross-references: EMBL:232701; NID:g475168; PID:g475169
C:Comment: This acid hydrolase catalyzes the cleavage of a wide variety of beta-glucuron C:Genetics:
A:Gene: uida
A:Map position: 36 min
C:Function:
A:Description: catalyzes hydrolysis of beta-D-glucuronoside to D-glucuronate and alcohol
A:Pathway: sacchar and glucose metabolism
C:Superfamily: beta-glucuronidase
C:Keywords: glycosidase; hydrolase

Query Match 29.3%; Score 879; DB 1; Length 603;
Best Local Similarity 33.8%; Pred. No. 3e-52; Indels 62; Gaps 11;
Matches 205; Conservative 110; Mismatches 229;

```

QY 1 MVRPQRNKRFFILINGVNNLEV-----TSKDRPIAVPGSMNEQYODLCY 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MLRPVETPRRIKXKLDGLAFSLDRENGCIDGRWMSALQESRAIAVPSFNDQPADDI 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 46 EE--GPFTKTFYVPKXLSQKHIRLYFAAVNTDCEVFLNGKRGVGNHIEYLPFVDTG 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 RNYAGVWYQREVFIPIKMGAGQRIVLRFDVTHYGKVVNNQBEVHEHGGYTFEADVTP 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 104 KYKSGEN-ELRVVVENRLKVGSPSKV---PSGTHTVGFGSPFPANDFPPYGGIIRP 159
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 YVIAGKSVRIIVCVNNELMWQITIPGMVITDEGKKQGSYF-----HDFNVAGIHRS 173
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 160 VLIETDHAIRIIDIWDTSESEBEKKLGKVKYKIEVSEBAQOEMTIKLGEEKKIRTSN 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 174 VMLYTPNTWVDITVTHVAQ---DCNIAVDWQ---VANGDVSVELRADQGVAVATG 227
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 220 RVEVEGFILNARFMSLEDPYLP---KVELEKDEYLDJGIRTISDEKRLYLNGKP 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 228 QGTSTGLQVNNPHLMQPGGYYELCVTKASQTECDIYLRGIRSVAAKGGQFLINHPK 287
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 276 VFLKGFSGKHEEPVVGQTFYPLMIDKDFMLKMINANSFRSHYPSSEWMLDLRLIGLT 335
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 288 FFFTFGRIEDADLRKGGFDNVLMDHDLMDVIGANSYRISHYPAEEMLMADBEHGIV 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 336 VIDEAPHVGI-----TRYHNPETQKIADNIRRMIDRKHNPVSVM 377
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 348 VIDETAAVGFNLSLIGFEAGNKKPKELYSEBAVNGETQQAHLQAIKELIARKNPSVVM 407
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 378 WSVANPEPSNHPDABSGFFKALYETANEMDRTRPVVMSMM--DAPDERTRVALYKPDIV 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 408 WSIANPEPTRPOGAREYFAPLAETARKLDPTRPIICVNVMPFCDATHTDISDL----FDVL 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

QY 436 CVNRYGVYIYQGRIBGLQALEKQIBELVARHHRKPIFVTEFGADAIAGIHYPDPQWSE 495
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 464 CUNRYGVYVOSGDLETRKEVLEKELAMQEKLPITITTEGVDTLGLHSMYTDWMS 523
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 496 EYQALVYKTRRLKDYIIGTHVMAFADKTPQNVARRPIINHGVFTTRDQPLVAVH 555
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 524 EYQCAMLMRYRVPFRVSAVVGEQVWNPADPATSGILRVQGNKKGIPTRRKPKSAFL 583
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 556 LRRLLMS 561
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 584 LQKRW 589
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 6
A26581
beta-glucuronidase (EC 3.2.1.31) precursor - human
C:Species: Homo sapiens (man)
C:Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 09-Jul-2004
C:Accession: A26581; A240337; A24983; A36538
R.Oshima, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly,
Proc. Natl. Acad. Sci. U.S.A. 84, 685-689, 1987
A:Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.
A:Reference number: A26581; MUID:87118233; PMID:3468507
A:Accession: A26581
A:Molecule type: mRNA
A:Residues: 1-651 <OSH>
A:Cross-references: UNIPROT:P08236; GB:M15182; NID:g183232; PIDN:AAA52561.1; PID:g183233
A:Experimental source: placenta
R.Shippley, J.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly
Genomics 10, 1009-1018, 1991
A:Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.
A:Reference number: A40337; MUID:92009900; PMID:1916806
A:Accession: A40337
A:Molecule type: DNA
A:Residues: 1-70 <SHI>
A:Cross-references: GB:M65002; NID:g183706; PIDN:AAA52622.1; PID:g183707
R.Gulise, K.S.; Korneluk, R.G.; Waye, J.; Lamhounah, A.M.; Quan, F.; Palmer, R.; Ganschow
Gene 34, 105-110, 1985
A:Reference number: A24983; MUID:85232043; PMID:3924735
A:Accession: A24983
A:Molecule type: mRNA
A:Residues: 520-585 <GUI>
A:Cross-references: GB:M10618; NID:g183704; PIDN:AAA52621.1; PID:g183705
R.Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeo, Y.; Yamada, S.; Yamada, Y.; Sasaki, T.;
Am. J. Hum. Genet. 48, 89-96, 1991
A:Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular het
A:Reference number: A36538; MUID:91090114; PMID:1702266
A:Accession: A36538
A:Molecule type: mRNA
A:Residues: 378-385, 616-621, 643-651 <TOM>
C:Genetics:
A:Gene: GDB:GUSB
A:Cross-references: GDB:120025; OMIM:253220
A:Map position: 7q22-7q22
C:Superfamily: beta-glucuronidase
C:Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
F.1-22/Domain: signal sequence #status predicted <SIG>
F.23-651/Product: beta-glucuronidase, placental #status predicted <MAT>

Query Match 28.8%; Score 865.5; DB 2; Length 651;
Best Local Similarity 34.5%; Pred. No. 2.8e-51;
Matches 215; Conservative 105; Mismatches 217; Indels 87; Gaps 21;

```

QY 1 MVRPQRNKRFFILINGVNNLEVTSKD-----RPI-----AVPGSMNEQ 39
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 27 MLYPQSPSRCKEIDGLMSFRADFSNRRRGFEQWYRRRLMSGPTVDMPPVSSFNDI 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 40 YOD--LCVEBGPFTYKTFYVPKXLSQ---KHILYFAAVNTDCEVFLNGKRGVGNHIEY 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 87 SQDMRLRFRVGVWVWEREVLLPERWTQDLRTTRVRLIGSASVAILWVNGVDTLHEHGGY 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 95 LPFVDTVYGVKXSG--ENEELR--VVVENRLKVGSPSKVP-----DSGTHTVGFGS 142
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

147 LPEADISNVQGPBLSRLRITIAINNTLT-----PTTLPBPGTIQYITDTSKPKGYF-- 200
QY 143 FPPANDFPPYGGIIRPVLIETFDHARILDIWDTSSEBEKKLGRVKKIEVSEAVGQ 202
DB 201 VQNTYDFPFYVAGLQSRVLLYTTPTTYIDITVLTTS---VEQDSGLVNVQISVGSNTL-F 256
QY 203 EMTIKLGEBEKKIRTSRPFVEGEFILENAPF-----SLEDP-VLYPLKYLE----- 249
DB 257 KLEVRLLDANKVAVNGTQGGQKLVPGVSLMWPYLMHRRPALYSLEVDLTQTSIGPV 316
QY 250 KQETVLDIGRTISWDEKRLYLNGKRPVFKGFKHEEPVLQGGTFYPLMKIDFNLLKMI 309
DB 317 SDYITLPVGRTAVALVTKSQPLNGKPPFYGVGVANKHEADLRGCGFMDLVLVDKDFNLLRL 376
QY 310 NANSFPTSHYPSYSEWMLDLADRLGILVIDEAPHVG-----TRYHNPETOKIAED 360
DB 377 GANAFTSHYPSYAEVWQMCDRYGIIVIDECPCVGLALPQFNNVSLHHH---MQVME 432
QY 361 NTRRMIDRKHNPESVIMWVSANEPESNHPABEGFFKALYETANEMDRTPRVNVVSMMDAP 420
DB 433 VVR---RDKNHPAVVWMSVANEPASHLESAGYLLKRVIAHTKSLDPSRPVTFVS---N 484
QY 421 DERTRDVALKYFPDIVCVNRYGVYIYQGRIEBGLQALEKDIEBLYARHRKPIFVTEFGAD 480
DB 485 SNVAADKGAAYVDVCLNSYTSWYHDYGHLELIQQLATQPEWYKTKYQKPIIQSEYGA 544
QY 481 AIAGIHYDPQWSESEYQAEVLEKTRILL--KQDIYIGTHVAFADFKTPQNVRRPIL 537
DB 545 TIAGFHQDPPLMFTEEYQKSLBQ-VYHGLDQKRKKVVGELIMNFDFTQESPTRVILG 603
QY 538 NHKGVFTRRQRPVAVHVR-RLM 560
DB 604 NKKGIYTRQRPKSAAPLLRERYW 627

RESULT 7
A85768
partial beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain ED
C/Spectes: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: A85768
R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
NATURE 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: A85768
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-368 <STO>
A/Cross-references: UNIPROT:Q8X671; GB:AE005174; MID:G12515602; PIDN:AAG56605.1; GSPDB:G
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: uidA_1

Query Match 23.0%; Score 691; DB 2; Length 368;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;
QY 230 NARFMSLEDPVLYPL-----KVELEKDEYTLDIGRTISWDEKRLYLNGKRPVFKGSGKE 285
DB 3 NPHLMQPGGYLYELVYTKASRTECDIYPLRVGIRSVAVKGEQFLNHRPFYFTGGRH 62
QY 286 EPPVILQGGTFYPLMKIDFNLLKVINANSFRTSHYPSYSEWMLDLADRLGILVIDEAPHVG 345
DB 63 DADLRKGGDNYVLMVHDHMLMDWIGANSYRTSHYPAEMLDMDADHGIIVIDETAAYGF 122
QY 346 -----TRYHNPETOKIAEDNIRRMIDRKHNPESVIMWVSANEPESN 387
DB 123 NLSLIGFPAQNKPKELYSEAVNGETQAHQAIKELIARDKNHPESVIMWVSANEPDR 182
QY 388 HPDABGFFKALYETANEMDRTPRVNVVSM--DAPERTRDVALKYFDIVCVNRYGVYI 445

DB 183 PGCAREYFAPLAEATRKLDPTRPITCVNVVMPCDADHTTISDL-----FDVLCLNRYGVYI 238
QY 446 YQGRIEBGLQALEKDIEBLYARHRKPIFVTEFGADAIAGIHYDPQWSESEYQAEVLEKT 505
DB 239 QSGDLETAKEVLEKELIAQOEKIQPIITTEGVDTLAIAGHSMTYDMSSEYQCAWLDWY 298
QY 506 IRLLEKDYIIGTHVAFADFKTPQNVRRPILNHKVPFRDRQPKLVAVHVRRLMS 561
DB 299 HRVFDVSAVVEQVWNVFADPATSQGLRVGKKGIFTRDRKPKSAAPLLQKRW 354

RESULT 8
D90919
beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain RMD 05092,
C/Spectes: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: D90919
R:Hayashi, T.; Makino, K.; Ohishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: D90919
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-370 <HAY>
A/Cross-references: UNIPROT:Q8X671; GB:BA000007; PIDN:BB35747.1; PID:G13361791; GSPDB:G
A/Experimental source: strain O157:H7, substrain RMD 050952
C/Genetics:
A/Gene: EGS2324

Query Match 23.0%; Score 691; DB 2; Length 370;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;
QY 230 NARFMSLEDPVLYPL-----KVELEKDEYTLDIGRTISWDEKRLYLNGKRPVFKGSGKE 285
DB 5 NPHLMQPGGYLYELVYTKASRTECDIYPLRVGIRSVAVKGEQFLNHRPFYFTGGRH 64
QY 286 EPPVILQGGTFYPLMKIDFNLLKVINANSFRTSHYPSYSEWMLDLADRLGILVIDEAPHVG 345
DB 65 DADLRKGGDNYVLMVHDHMLMDWIGANSYRTSHYPAEMLDMDADHGIIVIDETAAYGF 124
QY 346 -----TRYHNPETOKIAEDNIRRMIDRKHNPESVIMWVSANEPESN 387
DB 125 NLSLIGFPAQNKPKELYSEAVNGETQAHQAIKELIARDKNHPESVIMWVSANEPDR 184
QY 388 HPDABGFFKALYETANEMDRTPRVNVVSM--DAPERTRDVALKYFDIVCVNRYGVYI 445
DB 185 PGCAREYFAPLAEATRKLDPTRPITCVNVVMPCDADHTTISDL-----FDVLCLNRYGVYI 240
QY 446 YQGRIEBGLQALEKDIEBLYARHRKPIFVTEFGADAIAGIHYDPQWSESEYQAEVLEKT 505
DB 241 QSGDLETAKEVLEKELIAQOEKIQPIITTEGVDTLAIAGHSMTYDMSSEYQCAWLDWY 300
QY 506 IRLLEKDYIIGTHVAFADFKTPQNVRRPILNHKVPFRDRQPKLVAVHVRRLMS 561
DB 301 HRVFDVSAVVEQVWNVFADPATSQGLRVGKKGIFTRDRKPKSAAPLLQKRW 356

RESULT 9
D95842
probable beta-galactosidase (EC 3.2.1.23) [imported] - Sinorhizobium meliloti (strain 10
C/Spectes: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C/Accession: D95842
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Herman
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: D95842
A/Status: preliminary

A: Molecule type: DNA
A: Residues: 1-755 <KUR>
A: Cross-references: UNIPROT:O92XF7, GB:AL591985, PIDD:CA68404.1, PID:G15139876, GSPDB:G
A: Experimental source: strain 1021, megaplasmid pSymB
A: Galibbert, F.; Fhan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
pela, D.; Chailu, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A: Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure-
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A: Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A: Reference number: A96039; MUID:21368234; PMID:11474104
A: Contents: annotation
C: Geneticks:
A: Gene: lacZ1, SMD21655
A: Genome: plasmid
C: Superfamily: beta-galactosidase
A: Keywords: glycosidase; hydrolase

Query Match	16.2%	Score 487;	DB 2;	Length 755;
Best Local Similarity	25.7%	Pred. No. 2.5e-25;		
Matches 155;	Conservative 90;	Mismatches 193;	Indels 164;	Gaps 24;

```

OY 28 REPLAVESMNE-----OYODLCTHEEGFFYKTTFFYKXLSQKHRLRYAANTDCEVFLN 83
Db 30 QPISLFPNNAVELPBNFYFDERCYORA-FTYQRIVLAMPDFSGEVSILVYDAMADAVVYLN 88
OY 84 GEKXGENHIEYLPEFVDVYTKVSGENELRVVVENRKLKXGPPSVKVPSPGHTHTVFGSGF 143
Db 89 GEELVARDOTYTFPEARLTDRLLEGNNLTIVKID-----GSENPBIPFG----- 133
OY 144 PRANDEFPGYGIIRPVLIETDHAILEDIWDTSE-----SEPE----- 163
Db 134 --GRIDILTAAGYIRPVWLKVTDPPVSIANIKIETRDVSDTKAVSLRCDLSNPGSLFSG 191
OY 184 -----KLGKVKYKIEEVEEAVGQEMTITLGEEEKKITSNRPFVEGEPITLENAPFMSLD 238
Db 192 TISALLKONAAGEVLAVAGETTGOSLAFEM-----DG--LKGSLMIDID 234
OY 239 PVLVPLKVELE---KDEYTLDIGRTTISWDEKRLYLNGKPVFLKGFGEKHEPFLVQGT 294
Db 235 PVLVIVIEVELRTQOGSDSCFAAHGPRFATFTTEGRLNGRPLKIRGLNRHQSFFYVGYAM 294
OY 295 FYPLMTIDFPLMK--INANSFRTSHPYSEEMLDLADRLGITIVIDEAP--HVGITTYHY 350
Db 295 GRTAOEADADIMGRHCHNLVRTSHYPOSKWELDCDRIGLVEBELPGMWHIG----- 348
OY 351 NPETOKLAEDNIRRMIDRHKNHPSVIMWSV--ANEPESHNPDAEGFKALYETANEMDRT 409
Db 349 GEEKKQALINNVARMERMDNMHPSIYIMGVRLNESQDSH---DFYATTRLANELDPT 404
OY 410 PVVWVSMMDAPDERTBDVALKTFDIVCVNRYYGWYIYQARIEBGLQALEKDI---EELVA 466
Db 405 -----QTGGVR-----YITSEPLEDYVTWMDIFLNGEELPG 436
OY 467 RHKK-----PIFVYREGADALAGIHVDPPOMSEEL-QELVEKIRL 508
Db 437 ANRRRTALRPOOECTGLPRKVPPLITEFGG-----HMVPTKIYDOEORQEHNRHLE- 489
OY 509 LLLKKDY-----IISTHWAFADFEKTPONVARRPIILNHK-----GVFTRDQPLVA 553
Db 490 VLNAAVDDPDISAIGCMCFDNYT-----HKDGSGBRICYHGVMMDFEPRFAA 539
OY 554 HV 555
Db 540 YV 541

```

RESULT 10

beta-galactosidase (EC 3.2.1.23) - *Thermoaerobacterium thermosulfurigenes*
C:Species: *Thermoaerobacterium thermosulfurigenes*
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #next change 09-Jul-2004

C:Accession: J00275; J01963
R:Burchardt, G.; Bahl, H.
Gene 106, 13-19, 1991
A:Title: Cloning and analysis of the beta-galactosidase-encoding gene from *Clostridium* t
A:Reference number: J00275; PMID:9203905; PMID:1840542
A:Accession: J00275
A:Molecule type: DNA
A:Residues: 1-716 <BUR>

A:Cross-references: UNIPROT:P26225'; GB:MS7579; NID:g144838; PIDN:AAA23349.1; PID:g1448399
A:Experimental source: strain EMI
A:Accession: JI0963
A:Molecule type: protein
A:Residues: 1, 'X', 3-27 <BU2>

A;Molecule type: protein
A;Residues: 1,'X',3-27 <BU2>

C;Genetics:
A:Gene: lacZ

C;Superfamily: beta-galactosidase

F/389,429,462/Active site: Glu, Tyr, Glu #staac
C/keywords: glycosidase; homodimer; hydrolase

Query Match	16.1%;	Score 482;	DB 2;	Length 716;
Best Local Similarity	27.3%;	Pred. No. 5e-25;		
Matches 166;	Conservative 91;	Mismatches 201;	Indels 150;	Gaps 27;

```

0Y 30 IAVPSANEOYODLCYEEG-----PFTY-----KTFYVP 55
      : : : : :
      : : : : :
Db 5 IPINNWM---YFKADYEGYEKVDLRSFENVNLPHINILBYNNYFDEKMOYQIKSCYCKP 61
      : : : : :
0Y 60 KKL$OKH-----IRLYFAAVNTDCEVFLNGEKKVGNHIEYLPEE--VDVYGVKVGSGENJR 113
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 62 LHISEKYDXKTYIHFEGVMAVAQVYLLNGLYGHEKGGTTPPDIRIDEVYDMKKEINMLT 121
      : : : : :
0Y 114 VVENRLKVGSPSPKVPDPSGTHTVGFCGSPFPANDFPYPYGI IRPVLIETFDHAILDI 173
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 122 VVVDSTER-----SDIPPKG-----GQIDYLVGGIGYREVSGIYDVDFIKI 164
      : : : : :
0Y 174 WVDT--SESEPEKKL-----GKVYKIVSEVAVGOEMTIKIGBEEKKIRTSN 219
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 165 KVEITHGIDNESLSMLVHLENLHOGSNVFKVKINDKN--GKEVYK-----EFN 214
      : : : : :
0Y 220 RFVVG-----EPITEN---ARFWSLEDVYLYPLKYLE---KDEYTLIDIGIRIT$WDEK 267
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 215 TYLDVAVKDVSPFNIENTLKDILMDVDNPNLVEIKVGMKINNSDEVDNKGFRFAVFXPD 274
      : : : : :
0Y 268 RLVLNGKRVFLKGSGKHEEPVLGGCTYPLMIKOPNLIK--MINANSFTSHYVSEEMU 326
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 275 GFYNGRKLTKRGILNRHOSYPYVGAMPRRVQOEKDAELIKNELHINIVTSHYPO$KHPL 334
      : : : : :
0Y 327 DLABRLGILVDEAPHVIGITRYHVPEPOKTAEDNIRIMIDHKHNSPYIM$V--ANEPE 385
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 335 NKCBELGLLVEBELPG---WQYIGNSEMKKVAEQULREMITDMNHP$ILLGVRINBEO 391
      : : : : :
0Y 386 SNHPDAGEFFKALYETANEMDRTEPVVW$NMADADERTRVALKYFDIVCVNRYYGWYI 445
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 392 ---DDDAFYKMMNKIAHEIDPTROTGVRYI-----TNS$FLE--DV-----YT 430
      : : : : :
0Y 446 YQGRIEEEL--QALEEDIEELYARHKKPLFYTFEGADALAGIHYDPROM--SEEYQAEIYB 503
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 431 FNDPFIHDSINPLRQOEVETGLEHNVPLVYLYEYNG-----HMYPLTKRDNBERQ$EHCL 484
      : : : : :
0Y 504 KTI$RL---LLKKDYIIGTHVMAFADPKPOMVNR$RPLNKK-----GVFTRDROP 549
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 485 RHLRIQNSYI$DDSDISGAIGMCAFDYNT-----HKDGSGBRIGCYHGM$DMDFRLB 534
      : : : : :
0Y 550 KLVAVHVR 557
      : : : : :
      : : : : :
      : : : : :
      : : : : :
Db 535 KFASVYVK 542
      : : : : :
      : : : : :
      : : : : :
      : : : : :

```

RESULT 11

D86872
beta-galactosidase (EC 3.2.1.23) [imported] - Lactococcus lactis subsp. lactis (strain I
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence revision 23-Mar-2001 #ext change 03-Aug-2001

A:Reference number: A90981; MUID:84028567; PMID:6313347
A:Accession: A90981
A:Molecule type: DNA
A:Residues: 2-1024 <KAL>
A:Cross-references: UNIPROT:P00722; GB:V00296; NID:g41901; PIDN:CAA3573.1; PID:g197203
A:Note: translation of initiator Met is not shown
R:Powell, A.V.; Zablin, I.
J. Biol. Chem. 253, 5521-5525, 1978
A:Title: Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and
A:Reference number: A92233; MUID:78218239; PMID:97298
A:Accession: A92233
A:Molecule type: protein
A:Residues: 2-1024 <POW>
A:Note: This is the final paper in a series
R:Calof, M.P.; Miller, J.H.
Nature 285, 38-41, 1980
A:Title: Molecular consequences of deletion formation mediated by the transposon Tn9.
A:Reference number: A93224; MUID:80188189; PMID:6246435
A:Accession: A93224
A:Molecule type: DNA
A:Residues: 356-476 <CAL>
R:Rutshouser, E.C.; Richardson, J.P.
J. Mol. Biol. 208, 23-43, 1989
A:Title: Identification and characterization of transcription termination sites in the E
A:Reference number: S06878; MUID:89362462; PMID:2475637
A:Accession: S06878
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-147 <RUT>
A:Cross-references: EMBL:X16313; NID:g41903; PIDN:CAA34380.1; PID:g41904
R:Kryukov, N.N.; Petrov, N.A.; Karginov, V.A.; Vassiliev, S.K.
Bioorg. Khim. 6, 1735-1736, 1980
A:Title: Nucleotide sequence of a lambda-plac 5-1 DNA region coding for a COOH-terminal
A:Reference number: I41218
A:Accession: I41218
A:Status: translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 967-971, R', 973-1022, 'E', 1024 <MIK>
A:Cross-references: GB:M8327; NID:g146061; PIDN:AAA2835.1; PID:g146062
R:Jacobson, R.H.; Zhang, X.; Dubose, R.F.; Matthews, B.W.
submitted to the Brookhaven Protein Data Bank, July 1994
A:Reference number: A65162; PDB:1BGL
A:Contents: annotation; X-ray crystallography, 2.50 angstroms, residues 4-1024
R:Jacobson, R.H.; Zhang, X.D.; Dubose, R.F.; Matthews, B.W.
Nature 369, 761-766, 1994
A:Title: Three-dimensional structure of beta-galactosidase from E. coli.
A:Reference number: A58594; MUID:94277211; PMID:8008071
A:Contents: annotation; X-ray crystallography, 2.50 angstroms
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64761
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1024 <BLAT>
A:Cross-references: GB:A600041; GB:U00096; NID:g1786532; PIDN:AAC74447.1; PID:g1786539
A:Experimental source: strain K-12, substrain MG1655
R:Prentki, P.
Gene 122, 231-232, 1992
A:Title: Nucleotide sequence of the classical lacZ deletion delta M15.
A:Reference number: I40987; MUID:93083990; PMID:1339377
A:Accession: I40987
A:Molecule type: DNA
A:Residues: 1-11, 43-50 <RBG>
A:Cross-references: EMBL:X58252; NID:g40882; PIDN:CAA41206.1; PID:g40883
A:Genetics:
A:Gene: lacZ
A:Map position: 8 min
C:Complex: homotetramer
;Function:
;Description: catalyzes hydrolysis of lactose into galactose and glucose

C;Superfamily: beta-galactosidase
C;Keywords: glycosidase; homotetramer; hydrolase; magnesium
F/2-1024/Product: beta-galactosidase #status experimental <MT>
R/2-50/Region: alpha complementation
F/51-218/Domain: 1, jelly-roll beta-barrel #status predicted <DM1>
F/219-334/Domain: 2, fibronectin type-III fold #status predicted <DM2>
F/335-627/Domain: 3, distorted TIM barrel #status predicted <DM3>
F/628-737/Domain: 4, fibronectin type-III fold #status predicted <DM4>
F/738-1024/Domain: 5, anti-parallel beta-sandwich #status predicted <DM5>
F/147,419,462/Binding site: magnesium (Glu, His, Glu) #status experimental
F/462,504,538/Active site: Glu, Tyr, Glu #status predicted

Query Match 13.8%; Score 414; DB 1; Length 1024;
Best Local Similarity 28.4%; Pred. No. 3,7e+20;
Matches 132; Conservative 59; Mismatches 172; Indels 102; Gaps 16;

Dy LGVNVNLEVTISKDRDIAVGSMNE-----QYDDLCYE---EG 48
 ||| | | |
Db LNGEWRP--AWPRADPAVPESWLBCDLPRADVIVPSNMOMAGDAIYNNTVIPIVNPP 112
 ||| | | |
Oy PFT-----YKTFYPVKX-LSQKHRLYPEAVNTDCSEVFPLNGEKVENHIEYLPEVV 99
 ||| | | |
Db PFVPTENPTYGCYSLFFNNWDSWLGSGQRILIPFGVSAPFLMCGNRGVGGQSRLRPSEF 172
 ||| | | |
Oy DYTGYKVKGGENELRYVENRLKYGFSPSKYPDSSGHITVEFGSFPANPFPPYGIIIRP 159
 ||| | | |
Db DLSAFLRAGENRIALVMV-LRWSDGSY---LEDQ-----DWMRMSGIFRD 212
 ||| | | |
Oy VLIERTDHARILLDIWNDESSESREPKKLKVKYKIYESBAVQ----EMTIKKGESEKK 214
 ||| | | |
Db VSLLHKPTTOSDPFHVAATRFNDPS----RAVLAEVQCGLGRDYLRAYTVSLMQSBTG 267
 ||| | | |
Oy IRTSNRLFEEGERI-----LENARFMSLBEDPYLYLPKYLEKEDEXTL---- 255
 ||| | | |
Db VASGTARPFGGEIIDRGCYADRVTTLRNVENPRLMGAERINLVRAVELHTAOGTLLFAE 327
 ||| | | |
Oy 256 --DIGIRITSWDKERLYANGKFVFKGFGNHSEFPLQGOSTFYPLMIKDNLKWIANNS 313
 ||| | | |
Db ACDEVFRREVRLIENGCLLNIGKRPLLTRGVNRHEHNPHLHGQMVEDQTWMODILLMKQNNA 387
 ||| | | |
Oy FRTSYPRPVSEEWLDLADELGIIVIDEA---RH-VGITRRHYNPETOKIAMDNIRMIDR 368
 ||| | | |
Db VRCSHPVPHMLTYLTCDRYGLYYVDGANLETGMVPMNRRLTDPRMLPAMSERVTRYVROR 447
 ||| | | |
Oy 369 HKNHSVIMMSVANEP--ESNHPDAGEFPKAFUYETANMEMDRTAPV 411
 ||| | | |
Db 448 DRNHSVIVIMSLSGNESHGANK-----DALTRYIKSVDSPSRPV 485
 ||| | | |

RESULT 14
E90678
beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90678
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
GenBank Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno.
A;Reference number: A59629; MIDB:211562J1; PMID:11258796
A;Accession: E90678
A;Status: preliminary
A:Molecule type: DNA
C;Genetics:
A:Cross-references: UNIPROT:Q8X685; GB:BAA00007; PIDB:BAR33820.1; PID:g13359854; GSDB:D-G.
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Gene: EC0397
C;Superfamily: beta-galactosidase

Query Match 13.5%; Score 405.5; DB 2; Length 1024;
Best Local Similarity 25.2%; Pred. No. 1.4e+19;
Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;


```

0Y 15 ANGWNLEVTSKDRILAVGSMNE-----QYDLGCE--EG 48
Db 55 LNGEHW-QFWPFPABEAVPESWLECDLPEADDTVVVPSWOMHGADAPITVTPITVNP 112
0Y 49 PFT-----YKTFVVPKX-LSQKHITLYEAVNTDCEVFLNGEKGENHIEYLEPEV 99
Db 113 PFWPTENPTGCSYLTFNVDSESWLOEGQTRIIIFDGNNSAFHLMCMGRWVGYGQDSRLISEF 172
0Y 100 DVTGKVGSGENELRVVVENRLKVGCFPSKVPDSGHTVYGFSGFPFANFDFPPYGGIIRP 159
Db 173 DLSAFPLRGENRLAAVW-LRWSDGSY--LEDO-----DWMRSIGIFRD 212
0Y 160 VLIFFTDHARILIDIWDTSESEPEKKL---QKVAKVIESEAVQGEWITILGEEKKIR 216
Db 213 VSLHKPTQTQISDFHVATLFRNDPSPRAVLAEVQMGELRBL--KVTYSLMGETOYA 269
0Y 217 TSNRPFVGEPT-----LENAPFWSLEDPELYPLKYLEKEDEYTL----- 255
Db 270 SGTAPFGCEIIDERGGYADRVTLGLANVENPKLMSAIEIPNIYRAVELHTADGTLIEAEC 329
0Y 256 DIGARTISMDKRLIYLNGKPVFLGFGKEEPVULGQTFPLMIKQPNLLKMNANSFR 315
Db 330 DVGREVATIEGLLNLKGPULLINGVNRHEHPLHGOVMDQETVQODLLKONNFANVR 389
0Y 316 TSHRYPESEWDLADRLGLILVIDEA--PH-VGITRHYNPEYOKIAEDIRIMIRHK 370
Db 390 CSHTPNPRLWTLTLDRCRGLVYVDEANIEITHGMVPMNRLTDDPRMLPMSEKVTMTMGQRD 449
0Y 371 NHRBSVIMSVANEP--ESNHPDABGFPKALYETANENDRTRPV-----VMVS 415
Db 450 NHRPVIIWISLNEBESHGHNH-----DALRWIKSDVPSRPPVYEGGADTSATDIICP 502
0Y 416 MMDAPDERTRDVALKYPDIVCNRYUGWYIYQGRIBESGLALEKDIEBLVABHKKPIFYT 475
Db 503 MYARVDEBQPPNAPFKMSI-----KKWLSLPGEM-----RPLLLC 537
0Y 476 EFG--ADALAGIHYPDPOMFSEEOYA-----ELVEKTRILLKQDYIIGTH 519
Db 538 EYALAMGNSLGG-----FAKYQAQPRQYPRLOGFVMDLVNOS--LIKYDE--NGN 584
0Y 520 VMAE--ADFKTPQVVRPILNKHKGVFTRDRDP 549
Db 585 PWSAYGGDFGDTPNDRQFCMN--GLVFPADRP 614

```

RESULT 15
A:Accession: AB5529
beta-D-galactosidase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: AB5529
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
I.I.; L., J.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB5529
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1024 <STO>
A:Cross-references: UNIPROT:Q8X685; GB:AE005174; NID:q12513175; PIDN:AMG54693.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: lacZ
A:Superfamily: beta-galactosidase

Query Match	13.5%	Score 405.5;	DB 2;	Length 1024;
Best Local Similarity	25.2%;	Pred. No. 1.4e-19;		
Matches 159;	Conservative 88;	Mismatches 216;	Indels 169;	Gaps 26

```

15  LINGVNLVTSKDRPIAVPGSNVE-----QYDLCYE---EG 48
||| | : ||| || | | | | | | | | | | | | | |

```

Dh	55	LNGEM--QVMPFAPAEVAESWLECDLDPADTVVPSVMQNGHDAPIYTNVYPIVNP	112
Qy	49	PEF-----YKTFYPPKX--LSQKRLRYPAAVNDCEVFLNGEKYGENHIEYLPFEV	99
Dh	113	PFVPTENPTGCSLTFNWDSEWLDQGGRIIFDGNNSAFHLMCGRWGVGGDSRLISEF	172
Qy	100	DYVGYKSGENELRYVVENRLLKVGFPSKVPDSDGHTVGFSGFPFANPFPPFGIIRP	159
Dh	173	DLSAFLAENRLAAVNV--LRMSDSY---LEDQ-----DMKRSIGIFRD	212
Qy	160	VLIETDHAIRLIDVWDSISEPEKKL--GKVKKIEVSEAVGQEMTIKLGBEEKKIR	216
Dh	213	VSLHKKPTQISDFHVALTFNDFSRVALLEAVGUGELRDEL--RVTVSLMQGETQVA	269
Qy	217	TSNRPFEGEPI-----LMAFPMSLEDPYLRYLKVLEKDEYTL-----	255
Dh	270	SGTAPFGGELIDERGADRYTLGLAVNENPKLMSAELPNTRYAVVLELHTADGTLIEAEC	329
Qy	256	DIGIRTSIDEXERLYLNGKPVFLKQFGGHEEPVLAGOSTFYRLMKDFNLKMYNANSFR	315
Dh	330	DVGFREVRLENGLLILNGKPRLLIRVNHENHRLHGQVMEQITVQDILMKQNNPAAVR	389
Qy	316	TSHPYRSEEMLDLADRLGILVIDEA--PH--VGITRYHNPETOKIAEDNIRMIDRIK	370
Dh	390	CSHYPNHPLMYTLCDRYGLYVUDEANIEPTHGQVMPNNRLTDDPRLPAMSERVTRMQRDR	449
Qy	371	NHESVYMSVANER--ESNHPRABEFPKALYETANENDRTFRV-----VMVS	415
Dh	450	NHESVYIWSLGNESGHANH-----DALYRMKISVDPSPRYPVEGGADTSATIDICP	502
Qy	416	MMDAPERTRDVALKYFDIVCVNRYVGUYIYQGRILEGQLAEKDIEBELYARRKKPIFVT	475
Dh	503	MYARVDEDDPFAVVRKMSI-----KKRLSLRGEM-----RPLILC	557
Qy	476	EFG--ADAIAIHYDPQPMSEEVQA-----ELVEKTRILLKKDYIIIGTH	519
Dh	538	EYAHAMGNSLGG-----FAKUYQAFRQYPRLOGGFVMDLVDOS---LITKYE--NGN	584
Qy	520	VMAF--ADFKTPQVNRRLPLNHKKGYFTTRDQRP	549
Dh	585	PMSAYGGDGTDPNDQFCMN--GIVEADRTPE	614

Search completed: September 1, 2005, 19:33:26
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:14:24 ; Search time 177 Seconds

(without alignments)
1628.817 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MWRPQNKRRFTLLINGVNN.....TRDROPKLVAVHYRLRLMSV 563

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	299.9	99.9	563	2 Q9XOF2	Q9XOF2 thermotoga
2	101.1	33.7	570	2 Q97YI1	Q97YI1 sulfolobus
3	977.5	32.6	599	2 Q8X19	Q8X19 clostridium
4	972.5	32.4	599	2 Q8VNV4	Q8VNV4 clostridium
5	929	31.0	602	2 Q9AFV4	Q9AFV4 staphylococ
6	919.5	30.6	598	2 Q9AHJ8	Q9AHJ8 lactobacill
7	906	30.2	670	2 Q6NUG6	Q6NUG6 drosophila
8	905	30.2	656	2 Q9V8R0	Q9V8R0 drosophila
9	905	30.2	670	2 Q8MMB7	Q8MMB7 drosophila
10	902.5	30.1	648	1 BGLR_MOUSE	P12365 mus muscucu
11	899.5	30.0	648	1 Q6IR10	Q6IR10 mus muscucu
12	898	29.9	651	1 BGLR_FELCA	Q97524 felis silve
13	894.5	29.8	648	1 BGLR_RAT	P06760 ratu
14	888	29.6	651	1 BGLR_CANFA	Q18835 canis fami
15	882	29.4	603	2 Q93VY4	Q93VY4 arabidopsi
16	880	29.3	603	2 Q8F8A3	Q8F8A3 escherichia
17	879	29.3	603	1 BGLR_ECOLI	P05804 escherichia
18	879	29.3	603	2 Q9AHJ4	Q9AHJ4 escherichia
19	872.5	29.1	648	1 BGLR_CERAE	Q77695 cercopithe
20	865.5	28.8	651	1 BGLR_HUMAN	P08316 homo sapien
21	865	28.8	593	2 Q6A5C6	Q6A5C6 propionibac
22	862	28.7	628	2 Q95Q32	Q95Q32 caenorhabd
23	861.5	28.7	603	2 Q6WJ7	Q6WJ7 ruminococu
24	856.5	28.5	808	2 Q77PJ3	Q77PJ3 ratu
25	841.5	28.0	686	2 Q9V9T9	Q9V9T9 drosophila
26	821.5	27.5	599	2 Q8E6A6	Q8E6A6 streptococ
27	821.5	27.4	599	2 Q8E0N2	Q8E0N2 streptococ
28	811	27.0	567	2 Q7UCB6	Q7UCB6 shigella fl
29	809	27.0	567	2 Q83RC5	Q83RC5 shigella fl
30	728	24.3	660	2 Q7PZB2	Q7PZB2 anopheles g
31	691	23.0	368	2 Q8X671	Q8X671 escherichia

32	691	23.0	370	2 Q7ADL5	Q7ADL5 escherichia
33	531	17.7	459	2 Q8F8X0	Q8F8X0 corynebacte
34	467	16.2	755	2 Q92XF7	Q92XF7 rhizobium m
35	486.5	16.2	745	2 Q93IM0	Q93IM0 thermotoga
36	482	16.1	716	1 BGLR_THETU	P26257 thermotoga
37	481.5	16.0	998	1 BGLR_LACLA	Q48727 lactococcus
38	477.5	15.9	998	2 Q8VPU3	Q8VPU3 lactococcus
39	476.5	15.9	1085	2 Q8S250	Q8S250 thermotoga
40	472.5	15.7	996	2 Q87523	Q87523 lactococcus
41	460	15.3	743	1 BGLR_THETU	P77989 thermotoga
42	453	15.1	704	2 Q650P5	Q650P5 bacteroides
43	452.5	15.1	1084	1 BGLR_THEMEA	Q56307 bacteroides
44	441.5	14.7	695	2 Q64ZB7	Q64ZB7 bacteroides
45	431.5	14.4	682	2 Q8A9Q8	Q8A9Q8 bacteroides

ALIGNMENTS

RESULT 1	ID	Q9XOF2	PRELIMINARY;	PRT;	563 AA.
AC	Q9XOF2;				
DT	01-NOV-1999 (TrEMBLrel. 12, Created)				
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Beta-glucuronidase.				
GN	OrderedLocusNames=TM1062;				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_Taxid=23356;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109 / ATCC 43589;				
RX	MEDLINE=9287316; PubMed=10360571; DOI=10.1038/20601;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,				
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,				
RA	Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eilen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Frazer C.M.;				
RT	*Evidence for lateral gene transfer between Archaea and Bacteria from				
RT	genome sequence of Thermotoga maritima.;				
RL	Nature 399:323-329(1999).				
DR	EMBL; AE001766; AAD36143.1; -.				
DR	PIR; A72300; A72300.				
DR	HSSP; P08236; 1BHG.				
DR	TIGR; TM1062; -.				
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR008979; Gal bind like.				
DR	InterPro; IPR006101; Glyco_hydro_2.				
DR	InterPro; IPR006102; Glyco_hydro_219.				
DR	InterPro; IPR006104; Glyco_hydro_28B.				
DR	InterPro; IPR006103; Glyco_hydro_27IM.				
DR	Pfam; PF00703; Glyco_hydro_2; 1.				
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.				
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.				
DR	PRINTS; PR00132; GLYHDLASE2.				
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 563 AA; 65682 MW; 98C030B75D33B6C1 CRC64;				
Query Match					
Best local Similarity 99.9%; Score 2999; DB 2; Length 563;					
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1 MWRPQNKRRFTLLINGVNNLEVTSKDRPIAVGSMNEOYQDLCYBEGFTYKTTFYVVK 60				
DB	1 MWRPQNKRRFTLLINGVNNLEVTSKDRPIAVPGSMNEOYQDLCYBEGFTYKTTFYVVK 60				
QY	61 XLSQGHIRLYFAAVNTDCVFLNGEKVGGENHTEYLPEFVDYVGVKSGENELRVVVENRL 120				

```

Db      61 ELISKHRLRYFAVNTDCEVFLNGEKVGENHIEYLPEFVDVTKVYKSGENELRVVNERL 120
Qy      121 KVGSPSKVPDSGTHVGFSGSPRPNRDPFPYGGIIRPVLIETFDHARILDIWDTSS 180
Db      121 KVGSPSKVPDSGTHVGFSGSPRPNRDPFPYGGIIRPVLIETFDHARILDIWDTSS 180
Qy      181 EPEKLGKVKVIEVSEEAVALGEMTIKLGEEKKIRTSNRFVEGEFIIENAFMSLEDPY 240
Db      181 EPEKLGKVKVIEVSEEAVALGEMTIKLGEEKKIRTSNRFVEGEFIIENAFMSLEDPY 240
Qy      241 LYPFLVLEKEDYTLIDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGGGTYPYPMI 300
Db      241 LYPFLVLEKEDYTLIDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGGGTYPYPMI 300
Qy      301 KQNLKWTNANSFRTSHYPSYSEEWLADLRIGIIVIDEAPVGTIRHYNNEETOKIAD 360
Db      301 KQNLKWTNANSFRTSHYPSYSEEWLADLRIGIIVIDEAPVGTIRHYNNEETOKIAD 360
Qy      361 NIRMIDRKHNPVSIVMSVANEPESSNHPDAGCFKALYETANEMDRTPVVMVSMMDAP 420
Db      361 NIRMIDRKHNPVSIVMSVANEPESSNHPDAGCFKALYETANEMDRTPVVMVSMMDAP 420
Qy      421 DERTDVALKYPDIYCVNRYYGYIYQRIIEBGLQALEKDIIELYARHRKPIFTVEFGAD 480
Db      421 DERTDVALKYPDIYCVNRYYGYIYQRIIEBGLQALEKDIIELYARHRKPIFTVEFGAD 480
Qy      481 ATAGTHYDPPQMFSEEOALEVETKIRLLKKDYIIIGTHVMAFAPKPTQNVRRPILNHK 540
Db      481 ATAGTHYDPPQMFSEEOALEVETKIRLLKKDYIIIGTHVMAFAPKPTQNVRRPILNHK 540
Qy      541 GVFTDRQPKLVAVHLRLMSEV 563
Db      541 GVFTDRQPKLVAVHLRLMSEV 563

RESULT 2
Q97UI1 PRELIMINARY; PRT; 570 AA.
AC 097UI1;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Beta-glucuronidase (Gusb) (EC 3.2.1.31).
GN Name=gusb; OrderedlocusNames=SSO30136;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
NCBI_TaxID=2287;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.114222098;
RA She Q., Singh R.K., Contaloni F., Zivanovic Y., Alward G.,
RA De Moers A., Traubo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Koxera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006894; AAK43138.1; -.
DR PIR; C90485; C90485.
DR HSSP; P08236; 1BHG.
DR GO; GO:0004566; F:beta-glucuronidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal_bind_1like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TBM.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.

```

```

DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLYDRLASE2.
KW Complete proteome; Glycosidase; Hydrolase.
SQ SEQUENCE 570 AA; 66795 MW; DEB2FE08050AF189 CRC64;

Query Match      33.7%; Score 1011; DB 2; Length 570;
Beet Local Similarity 39.2%; Pred. No. 5,2e-58;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

Qy      15 LQVWNLLEVTSKDRP-----IAPGSMNQYODLCIEEGPFTYKTFYYPK 60
Db      11 LQGFKKFKIDENNGEENGWYKGLSESDIIVPASMNQNPMDQFSGIAPVQKDLFVSN 70
Qy      61 XLSOKHRLRYFAVNTDCEVFLNGEKVGENHIEYLPEFVDVTKVYKSGENELRVVNERL 120
Db      71 DNGRKAMVVEGAGYITKLMINGEYGTTHSGFTQGFPIKLKV---NEENKLV---V 123
Qy      121 KVGSPSKVPDSGTHVGFSGSPRPNRDPFPYGGIIRPVLIETFDHARILDIW 174
Db      124 KIDNTPSPY-----NLPPARDLNNAAFDFNYGGIHRVYIEFVDECHVEDIT 171
Qy      175 VDTSSSEPEKLGKVKVIEVSEEAVALGEMTIKLGEEKKI---RTSNRFVEGEFIIEN 230
Db      172 VYT-----KSYGHLKVEI-LSECNQRPSSLKFLVDKGRVILNESSNEVEFKD--VNN 222
Qy      231 ARFMSLEDPVLYPLKVELE-----KDEYTLIDIGIRTIISWDEKRLYLNGKPVFLKGFGKHE 285
Db      223 VIPMSDPNPLYTLIVENHYVGNLKDYSYERIGRDEVVKOKIYLNKGRPIFLKFGFHE 282
Qy      286 EPPVLGGTFFPLMIKQNLKWTNANSFRTSHYPSYSEEWLADLRIGIIVIDEAP--HV 343
Db      283 DFLIGKFTYGAIVLRDLYLRKIGANSFRTSHYPSYSENEHLDLADMEGFLVILBPLCY 342
Qy      344 GTRHYHNPETQKI-----AEDNIRMIIDRKHNPVSIVMSVANEPESSNHPDAGCF 394
Db      343 NISRVMSQEEIAKMGVDYKFEKVRDTIKEMIIRQKHNPVSIVMSVANEPESSNHPDAGCF 402
Qy      395 FKALYETANEMDRTPVVMVSMMDAPDERTDVALKYPDIYCVNRYYGYIYQRIIEBGL 454
Db      403 IRREVELFKSIDSSRPVTFAS-----HRSVADLALVYDVLSINTYHGHYTWGIDISGV 457
Qy      455 QALEKDIIELYARH-RKPIFTVEFGADAIAGIHYDPPQMFSEEOALEVETKIRLLKKD 513
Db      458 KVAIAIELEIRHKFKPEKPIIITFEADAIYGLHSDPPQMFSEEOALEVETKIRLLKKD 517
Qy      514 YIIGHVMAFADFKTPQNVRRPILNHKGVFTDRQPKLVAVHLRLM 560
Db      518 YIVGFHINVFADFRTPQNPSTRILNRKGIFTDRQPKLVAAVVEELF 564

RESULT 3
Q8XP19 PRELIMINARY; PRT; 599 AA.
ID Q8XP19;
AC Q8XP19;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Beta-glucuronidase.
GN Name=BgIR; OrderedlocusNames=CPB0147;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
NCBI_TaxID=1502;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=13;
RX MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shintzu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79853.1; -.

```

DR HSP; P08236; 1BHG.
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR Pfam; PF00703; Glyco_hydro_2_1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; UNKNOWN_1.
 DR Complete proteome.
 KM SEQUENCE 599 AA; 68729 MW; 1AE3393869DAEF3B CRC64;
 SQ
 Query Match 32.6%; Score 977.5; DB 2; Length 599;
 Best Local Similarity 36.9%; Pred. No. 8,7e-56;
 Matches 227; Conservative 102; Mismatches 210; Indels 77; Gaps 16;
 QY 1 MVRPQNKRRFILLINGVNLVETSKDR-----PIAVGSMNE--QYQDLC 44
 DB 1 MLYPIITTESRQILIDLSGIMWFKLNEGNGLTLEELSKAPLEDITIMAVPSSYNLDVLESQEV 60
 QY 45 YEEGPTYKTFYVPKLSQKHRLYFAAVNTDCEVFLNGEKGENHIEYLPEVDVTK 104
 DB 61 DHVGWVWYERNFTIPRTLNERIVLRFSGATHEAKYVINGELLVHKGGFTEPEAEINDL 120
 QY 105 VSGENELRVVENRLKVGFPSPGSGTHTVGFFGSP-----PANDFPPYG 155
 DB 121 LVSGDNRLLTVAVNN-----IIDETTLPVGLVKEVEVDGKVIKSNVDFPNYAG 170
 QY 156 IIRPVLIETDARILIDWDTSESSEPKGLGKVKYIVSEBAVQ-EMTIKLGEEKK 214
 DB 171 IHRPVKIYTPKSYIEDITIVDFKENN--GYVNVV-----QAVGKNIKVTIIDENN 223
 QY 215 IRTSNRFVGEFLENAFMSLEDPLYPLKYLEKDEYTLD-----IGRTISMDEKRL 269
 DB 224 IYABEGGKGGKLTINNVMLEPMNAVLYLKYLELDDDEIITDYEFEFGVRIVEVDGKF 283
 QY 270 YLNGKRVPLKGFGEHEEPVLGGTFYPLMKIDFNLLKWINANSFRTSHYPYSEWLDLA 329
 DB 284 LINNPFYKGFGEHSDSYVNGRGINEAINIKDFNLKMGANSFRTSHYPYSEIMRLA 343
 QY 330 DRLGLIVIDEAPHVGITRYHYN-----PE-----TOKIAEDNIRRMIDRHN 371
 DB 344 DREGIVVIDETPAVGL--HLNFMATGFGDAPKRDWTKEIGTKEAHERILRLVSRDN 400
 QY 372 HPSVIMWSVANEPSNHPDABGFKALYETANEMD-RTRPVVMS-WMDAPDE-RTRDVA 428
 DB 401 HCVVMWSVANEPSDSEGAKEYFEPLIKLTKELDPOKRPVVTVVYLMSTPRCKVDIV 460
 QY 429 LKFPDIVCNRYGYWYIYQRIEGLQALEKOIEELVYARHK-PIFVTEFGADATAGHY 487
 DB 461 ----DVLCLNRYGYWYVAGDLEBAKRMLEDLKGMEBCPKTPIMFTEYGADTVAGLHD 516
 QY 488 DPPQMFSEYQAEVLVEKTRILLLKKDYIIGTHVMAFADFKTPQNVARRPILNHKVFTDR 547
 DB 517 TVPVMTTEBYQVEYYKANHEVMDCKNPFGEQVWNAFADPATSGIIRVQGNKKGIPTRR 576
 QY 548 QPKLVAVLRLMSEV 563
 DB 577 KPKMIAHSRERWTNI 592
 RESULT 4
 Q8VNV4 PRELIMINARY; PRT; 599 AA.
 AC 08VNV4;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Beta-glucuronidase.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OK NCBI_Taxid=1502;

RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=13;
 RX MEDLINE=21945361; PubMed=11948145;
 RA DOI=10.1128/JB.184.9.2333-2343.2002;
 RA Briolat V., Reyssat G.;
 RT "Identification of Clostridium perfringens genes involved in the
 RT adaptive response to oxidative stress.";
 RL J. Bacteriol. 184:2333-2343(2002).
 DR EMBL; AJ20784; CAD12654.1; .
 DR HSP; P08236; 1BHG.
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR Pfam; PF00703; Glyco_hydro_2_1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; UNKNOWN_1.
 DR SEQUENCE 599 AA; 68715 MW; BB5336CA4AAED48 CRC64;
 SQ
 Query Match 32.4%; Score 972.5; DB 2; Length 599;
 Best Local Similarity 36.7%; Pred. No. 1.9e-55;
 Matches 226; Conservative 102; Mismatches 211; Indels 77; Gaps 16;
 QY 1 MVRPQNKRRFILLINGVNLVETSKDR-----PIAVGSMNE--QYQDLC 44
 DB 1 MLYPIITTESRQILIDLSGIMWFKLNEGNGLTLEELSKAPLEDITIMAVPSSYNLDVLESQEV 60
 QY 45 YEEGPTYKTFYVPKLSQKHRLYFAAVNTDCEVFLNGEKGENHIEYLPEVDVTK 104
 DB 61 DHVGWVWYERNFTIPRTLNERIVLRFSGATHEAKYVINGELLVHKGGFTEPEAEINDL 120
 QY 105 VSGENELRVVENRLKVGFPSPGSGTHTVGFFGSP-----PANDFPPYG 155
 DB 121 LVSGDNRLLTVAVNN-----IIDETTLPVGLVKEVEVDGKVIKSNVDFPNYAG 170
 QY 156 IIRPVLIETDARILIDWDTSESSEPKGLGKVKYIVSEBAVQ-EMTIKLGEEKK 214
 DB 171 IHRPVKIYTPKSYIEDITIVDFKENN--GYVNVV-----QAVGKNIKVTIIDENN 223
 QY 215 IRTSNRFVGEFLENAFMSLEDPLYPLKYLEKDEYTLD-----IGRTISMDEKRL 269
 DB 224 IYABEGGKGGKLTINNVMLEPMNAVLYLKYLELDDDEIITDYEFEFGVRIVEVDGKF 283
 QY 270 YLNGKRVPLKGFGEHEEPVLGGTFYPLMKIDFNLLKWINANSFRTSHYPYSEWLDLA 329
 DB 284 LINNPFYKGFGEHSDSYVNGRGINEAINIKDFNLKMGANSFRTSHYPYSEIMRLA 343
 QY 330 DRLGLIVIDEAPHVGITRYHYN-----PE-----TOKIAEDNIRRMIDRHN 371
 DB 344 DREGIVVIDETPAVGL--HLNFMATGFGDAPKRDWTKEIGTKEAHERILRLVSRDN 400
 QY 372 HPSVIMWSVANEPSNHPDABGFKALYETANEMD-RTRPVVMS-WMDAPDE-RTRDVA 428
 DB 401 HCVVMWSVANEPSDSEGAKEYFEPLIKLTKELDPOKRPVVTVVYLMSTPRCKVDIV 460
 QY 429 LKFPDIVCNRYGYWYIYQRIEGLQALEKOIEELVYARHK-PIFVTEFGADATAGHY 487
 DB 461 ----DVLCLNRYGYWYVAGDLEBAKRMLEDLKGMEBCPKTPIMFTEYGADTVAGLHD 516
 QY 488 DPPQMFSEYQAEVLVEKTRILLLKKDYIIGTHVMAFADFKTPQNVARRPILNHKVFTDR 547
 DB 517 TVPVMTTEBYQVEYYKANHEVMDCKNPFGEQVWNAFADPATSGIIRVQGNKKGIPTRR 576
 QY 548 QPKLVAVLRLMSEV 563
 DB 577 KPKMIAHSRERWTNI 592
 RESULT 5
 Q9AFA2 PRELIMINARY; PRT; 602 AA.
 ID Q9AFA2

AC O9AFA2; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Beta-glucuronidase.
 OS Staphylococcus sp. RLH1.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 NCBI_TaxID=156489;
 RN NCBI_TaxID=156489;
 RP SEQUENCE FROM N.A.
 RC STRAIN=RLH1.
 RA Jefferson R.A., Keese P.K.;
 RL Submitted (JEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RA Killian A.;
 RL Submitted (JEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354044; AKK29422.1; -.
 DR HSP; P08236; 1BHG.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_2ig.
 DR InterPro; IPR006103; Glyco_hydro_2SB.
 DR InterPro; IPR006104; Glyco_hydro_2TIM.
 DR Pfam; PF00703; Glyco_hydro_2_1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 SQ SEQUENCE 602 AA; 68701 MW; 312AFCD1634D577 CRC64;

Query Match 31.0%; Score 929; DB 2; Length 602;
 Best Local Similarity 35.6%; Pred. No. 1.3e-52;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

QY 1 MVRPQNKRRFLLILNGVNNLEV-----TSKDRINAPGSMNE--QYDLC 44
DB 1 MVRPQNKRRFLLILNGVNNLEV-----TSKDRINAPGSMNE--QYDLC 44
QY 1 MVRPQNKRRFLLILNGVNNLEV-----TSKDRINAPGSMNE--QYDLC 44
DB 1 MVRPQNKRRFLLILNGVNNLEV-----TSKDRINAPGSMNE--QYDLC 44
QY 45 YEBGPFTYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104
DB 45 YEBGPFTYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104
QY 61 NHIGVWYREFVPAVKDQRLVLRFGSATHKALYYVANGELVHKGGLPFPEAIIINS 120
DB 61 NHIGVWYREFVPAVKDQRLVLRFGSATHKALYYVANGELVHKGGLPFPEAIIINS 120
QY 105 VASGENELRVVENRLKVGGFPSKVPDSGTHVGFPGS-----PPANFDFPYG 154
DB 105 VASGENELRVVENRLKVGGFPSKVPDSGTHVGFPGS-----PPANFDFPYG 154
QY 121 LRDMGRVTVVAVDNLI-----DSTLPVGLYSERHEBGLKVINRKNPFDFPNYA 170
DB 121 LRDMGRVTVVAVDNLI-----DSTLPVGLYSERHEBGLKVINRKNPFDFPNYA 170
QY 155 GIIRPULIEFTDARLIDIMVDTSESEPEKKLGKVKKLEVESEAVGQEMTIKLG--EE 212
DB 155 GIIRPULIEFTDARLIDIMVDTSESEPEKKLGKVKKLEVESEAVGQEMTIKLG--EE 212
QY 171 GLHRPXYKITYTPTVEDISVTDENGPR--TGTVTYVDFO---GKAETKVSVDDE 223
DB 171 GLHRPXYKITYTPTVEDISVTDENGPR--TGTVTYVDFO---GKAETKVSVDDE 223
QY 213 KKIIRTSNRVEGEFLEINARFMSLEDPYLPKLVLEKDEYTLDI-----GIKRTISMDK 267
DB 213 KKIIRTSNRVEGEFLEINARFMSLEDPYLPKLVLEKDEYTLDI-----GIKRTISMDK 267
QY 224 GKVAATBEGISGVNELPNVILMEPLNTYLYQIKVELVNDGLITIDVBEFPGRTYEVNDG 283
DB 224 GKVAATBEGISGVNELPNVILMEPLNTYLYQIKVELVNDGLITIDVBEFPGRTYEVNDG 283
QY 268 RLYLNGKVPVPLKGFGEHEEPVVGQTFVPLMIKIDFNLLKWINANSFRTSHYPSYSEMD 327
DB 268 RLYLNGKVPVPLKGFGEHEEPVVGQTFVPLMIKIDFNLLKWINANSFRTSHYPSYSEMD 327
QY 284 KFLINNKPFYKFGFGHEDTPIINGRGNFNASVMDENILKIGANSFRTAHYPSSELMR 343
DB 284 KFLINNKPFYKFGFGHEDTPIINGRGNFNASVMDENILKIGANSFRTAHYPSSELMR 343
QY 328 LADRIGILVIDEAPHGIRRYHN-----PETQKIA-----EDNIRMDIR 368
DB 328 LADRIGILVIDEAPHGIRRYHN-----PETQKIA-----EDNIRMDIR 368
QY 344 LADREGLVVIDETPAVG--HINFMATTGLGGSRSVSTWEKIRTFEHHQDVLRLYSR 400
DB 344 LADREGLVVIDETPAVG--HINFMATTGLGGSRSVSTWEKIRTFEHHQDVLRLYSR 400
QY 369 HKNHPSVIMSVANEESENHPDAEGFKALYETANEMD--RTRPVVWVWMDAPDESTRV 427
DB 369 HKNHPSVIMSVANEESENHPDAEGFKALYETANEMD--RTRPVVWVWMDAPDESTRV 427
QY 401 DKNHPBVVWVWMSIANEAATEBEGAYEFKPLVELTKLDPOKRPVTLVLFVMAITPE--TDK 458
DB 401 DKNHPBVVWVWMSIANEAATEBEGAYEFKPLVELTKLDPOKRPVTLVLFVMAITPE--TDK 458
QY 428 ALKYPDIVCVNRYGYIYQGRIEBGLQALEKDIEELVYARH--KPIFTEFGADATA 483
DB 428 ALKYPDIVCVNRYGYIYQGRIEBGLQALEKDIEELVYARH--KPIFTEFGADATA 483

```

DB 459 VAEIIDLIALNRNGMYFDGDLBAKVHLR---QEFHAMNKRCSGPKIMITEYGADTVA 515
 QY 484 GIHYRPPOMFSEYQALVEKIRILLKKDYIIGHWAFAFDPKPPVNRRLTNHKGVF 543
 DB 516 GFHDIDPWFTEEYVEYYQANHVFDEFENFVGQANFADFATISQGVMRVQGNKKGVF 575
 QY 544 TRDROPKLVAHYRLRLMSEV 563
 DB 576 TRDRPKLAAHYFRERMTNI 595

RESULT 6
 ID O9AHU8 PRELIMINARY; PRT; 598 AA.

AC O9AHU8; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Beta-glucuronidase.
 GN Name=gusA;
 OS Lactobacillus gasseri.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 NCBI_TaxID=1596;
 RN NCBI_TaxID=1596;
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADH.
 RX MEDLINE=21141841; PubMed=11229918;
 RX DOI=10.1128/AEM.67.3.1253-1261.2001;
 RA Russell W.M., Kleenhammer T.R.;
 RT "Identification and cloning of gusA, encoding a new beta-glucuronidase
 from Lactobacillus gasseri ADH."
 RL Appl. Environ. Microbiol. 67:1253-1261(2001).
 DR EMBL; AF305888; AK07836.1; -.
 DR HSP; P08236; 1BHG.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_2ig.
 DR InterPro; IPR006103; Glyco_hydro_2SB.
 DR InterPro; IPR006104; Glyco_hydro_2TIM.
 DR Pfam; PF00703; Glyco_hydro_2_1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 SQ SEQUENCE 598 AA; 69762 MW; 5398F06082DD887D CRC64;

Query Match 30.6%; Score 919.5; DB 2; Length 598;
 Best Local Similarity 34.1%; Pred. No. 5.6e-52;
 Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;

```

QY 4 PQRNKKRFTLLILNGVNNLEVT-----SKDRP-----IAPGSMNE--QYDLCYEE 47
DB 4 PQRNKKRFTLLILNGVNNLEVT-----SKDRP-----IAPGSMNE--QYDLCYEE 47
QY 7 PQRNKKRFTLLILNGVNNLEVT-----SKDRP-----IAPGSMNE--QYDLCYEE 47
DB 7 PQRNKKRFTLLILNGVNNLEVT-----SKDRP-----IAPGSMNE--QYDLCYEE 47
QY 48 GPFTYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGKYS 107
DB 48 GPFTYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGKYS 107
QY 67 GDFWQKDFPFPSPFLKKKELYIRFGSVTHRAKVPFINGHEVQHEGGLPFOVKISNYINY 126
DB 67 GDFWQKDFPFPSPFLKKKELYIRFGSVTHRAKVPFINGHEVQHEGGLPFOVKISNYINY 126
QY 108 GE-NELRVVENRLKVGGF--SKVPSGSHVYGFPSFPANDFPPYGIIRPVLI 163
DB 108 GE-NELRVVENRLKVGGF--SKVPSGSHVYGFPSFPANDFPPYGIIRPVLI 163
QY 127 DQTRVTVLVNNEISEKAIPTGTEIILDNG-----KLAQPYDFFNYSGIMNVWL 179
DB 127 DQTRVTVLVNNEISEKAIPTGTEIILDNG-----KLAQPYDFFNYSGIMNVWL 179
QY 164 FTDHARLIDIMVDTSESEPEKKLGKVKKLEVESEAVGQ--EMTIKLGEEKKIRTSNRFV 222
DB 164 FTDHARLIDIMVDTSESEPEKKLGKVKKLEVESEAVGQ--EMTIKLGEEKKIRTSNRFV 222
QY 180 ALPQSQINPFLN-----YQLANNKATITTYNIEANNNAEFKVTLPFNGQEVACATSKN 232
DB 180 ALPQSQINPFLN-----YQLANNKATITTYNIEANNNAEFKVTLPFNGQEVACATSKN 232
QY 223 EGEFLLENARFWSLEDPYLPKLVLEK-----DRTYLDIGIRTSWDEKRLYLNGKVPF 277
DB 223 EGEFLLENARFWSLEDPYLPKLVLEK-----DRTYLDIGIRTSWDEKRLYLNGKVPF 277
QY 233 TSSLTIKPHLMSVNDPYSYKIKTEMLDGKTVDEYTDKIGIRTVKIVNDKILLNHPDIY 292
DB 233 TSSLTIKPHLMSVNDPYSYKIKTEMLDGKTVDEYTDKIGIRTVKIVNDKILLNHPDIY 292

```

Qy	278	LKFGFGHEEPVYGGCTFPVIMKDNKLKWTIANSPRSHYPSSEMDLADRLGIVY	337
Db	293	LKFGFGHEEPVNLGKAVNESIIKRDYECMKIGANCRRSHHYAAEWOYAKYGLII	3522
Qy	338	DEAPHVIGIRRYHN-----PETOKIAEDNIRRMIDRKHNPSTVIMS	379
Db	353	DEVPANGLNRSLTNFLVNTNSNSQSHFPASKTVTELKKNHGELEKEMIDDRHPSTVIMS	412
Qy	380	VANEPSSNHPDABGEFFKALYETIANEMD-RTRPVVMVSMMDAPEDKTRDVALKYFDIVCVN	438
Db	413	LENPEPSTTQESYDYFKDIFAFARKLDPQNRPTGTGLVMGSGPK--VDKLHPCLDDEVCLN	470
Qy	439	RYGWTIYCG-RIEBSGLALEKDIIEELY-ARHKRPIFTYFQADLTAAGIHYPDPQFSE	496
Db	471	RIYGMVAGGPEELVNAKKMLDELDQWNLKLNKPPVFTEFGADTSSSHRLPDEWMSOE	530
Qy	497	YQAEIVKERTIRLLKKDYIIGTHVMAFADFKTEPQNRRLPLNKGGVTRDROPKLVAHYL	556
Db	531	YQNEYVQWYFDLPFKKYPFCIGELVWVFAFPKISEGIMRGANDKGLFTTRDREKDIAPFL	590
Qy	557	RLMSEV 563	
Db	591	KKRWOL 597	
RESULT 7			
Q6NL66		PRELIMINARY; PRT; 670 AA.	
ID	Q6NL66		
AC	Q6NL66		
DT	05-JUL-2004	(TREMBLrel. 27, Created)	
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)	
DE	RELS795D.		
GN	Name=CG15117;		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkely;		
RA	Strapleon M., Carlson J., Chavez C., Frieze E., George R., Paclleb J.,		
RL	Parf S., Wan K., Yu C., Rubin G.M., Gelinker S.;		
RL	Submitted (Apr-2004) to the EMBL/genbank/DBD databases.		
DR	EMBL; BT012475; AAS93746.1; ..		
DR	GO; GO:0004553; F-hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.		
DR	GO; GO:0005975; P-carbohydrate metabolism; IEA.		
DR	InterPro; IPR008979; Gal_bind_like.		
DR	InterPro; IPR006101; Glyco_hydro_2.		
DR	InterPro; IPR006102; Glyco_hydro_2ig.		
DR	InterPro; IPR006103; Glyco_hydro_2SB.		
DR	InterPro; IPR006025; Pept_M_zn_BS.		
DR	Pfam; PF00703; Glyco_hydro_2; I.		
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.		
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.		
DR	PRINTS; PR00132; GLHYDRALASE2.		
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.		
DR	PROSITE; PS00608; GLYCOSYL_HYDROL_F2_C; 1.		
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN; 1.		
QO	SEQUENCE 670 AA; 77050 MW; 00AE0E67AE1D9E8C CRC64;		
Query Match 30.2%; Score 906; DB 2; Length 670;			
Best Local Similarity 34.7%; Pred. No. 5e-51;			
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17			
Qy	1	MVRPQRKKRPFLLINGVNT-----EYTSKDR---PIAVPSMWEQ 39	
Db	48	MLYPSSETRREVRSDGIMNFRSDQANPTQGRDEWYAKELSKSRPTIMPVPSYNDI 107	
Qy	40	YQD-LCYEGPPTYTTTFVVPKLSQ-KHIRLYFAAVNTDCEVPLNGENYGENHLEYLPF 97	

Db	108	TTDNLRDHVGITWYTRKFFVPSMSKDOIWLRFSGVHYEAYVWINGQVYKHEMGLPE	167		
Qy	98	EVDVYGVKYSG-ENELRVVENRLKVGFP----	SKVPDSCGTHTVGFGSPFPAHPFP	1522	
Db	168	EAEVTDLLSYGAENRITWVCNALLQTVTPQGRITFEPVDDGMTI-----	VGSYTFDDFN	2222	
Qy	153	YGGIIRPVLIEETDARILIDIVDTSSESEPEKKLGKVKKIEVSEAVGO----	EMTIK	207	
Db	223	YAGIHRSHVHLTYTPPTFIEEVETNLSK-DATIGEVFSVSGSAAEADNLVIOAN	281		
Qy	208	IGEEKKI---RTSRPFVEGEFIIENARFW-----	SLEDPYLYPLKVELEK-----	DEV	253
Db	282	LYDKGCIILVANAISQKGGKQVNPVKWMPYLMHSEGCYLYQLEIKLLATNDELDDY	341		
Qy	254	TLDIGIRITSMDEKLYLNGKPEVPLKGFGEKHEEPVLCQGTFTYPLMIKDFNLKWINANS	313		
Db	342	RLKVGIRFTLSWMSQGFPLNGKRVYRGFRHHSDIRGGLDNALMVRPNLLKMGANA	401		
Qy	314	FRTSHYPIVSEEMLADRLGLIIVDAPHVGIITRYHNDETOKIADNIRIMIDRKHNP	373		
Db	402	YRTHSHYPYSEESMOQFADEHGIMIIDECSEVDTE--NFSQELGKHKSILEQILHRDRNP	459		
Qy	374	SVIMSVANPEPSNHPDAEGFEKALYETANEMDRPVYVMSMDAPDERTDVALKYPD	433		
Db	460	SVMMVSIANEPTGVSADSYFELVANFTRSJDKTRPIAALAV---SNTODKARSJD	515		
Qy	434	IVCVNRYYGWYLYQGRIEBGLQALKEIDELYA---RHRKPIFTEFGADALAGIHYDP	490		
Db	516	IISFRVNAVMSNMRLD---MIONVIDEALANKRKNYKPIIMSEYADTIEGLHMOA	572		
Qy	491	QMFSESEYALVEKTIRL---LLKQDYIIIGHVMAPADFKTPQNVRRPILNKHGVETDR	547		
Db	573	YWMSEEFOTEVFSRHFKAFFDELRRKGWFIGEFVMNFAFKTAQSYTRVCGNKKGVETRAR	632		
Qy	548	QPKLVAYHLR	558		
Db	633	QPKAAHLRLK	643		

DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE CG15117-PB.
 GN ORFNames=CG15117;
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Blandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazew R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abil J.F., Achpani A., An H.J., Andrews-Pfannoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Iobagiam C.,
 RA Jalali M., Kalish P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulig D., Lai Z.,
 RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palenik K., Remington K., Saunders R.D., Schebler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson R., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Waasman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zheng F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champs M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodegryn E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT *melanogaster* euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]

RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG Submitted (MAR-2004) to the EMBL/Genbank/DBJ databases.
 RL EMBL; AF003797; AF006434.1; -;
 DR HSSP; P08236; IBHG.
 DR FlyBase; Fg00034417; CG15117.
 DR GO; GO:0004553; P:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_2ig.
 DR InterPro; IPR006104; Glyco_hydro_2SB.
 DR InterPro; IPR006103; Glyco_hydro_2TIM.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR Pfam; PF00703; Glyco_hydro_2_I.
 DR Pfam; PF02836; Glyco_hydro_2_C_1.
 DR Pfam; PF02837; Glyco_hydro_2_N_1.
 DR PRINTS; PR00132; GLHYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
 SQ SEQUENCE 670 AA; 77036 MW; 7A9F23351E458C90 CRC64;
 Query Match 30.2%; Score 905; DB 2; Length 670;
 Best local Similarity 34.7%; Pred. No. 5 8e-51;
 Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;
 1 MVRPQNRKFRILLINGVNT-----EYTSKDR---PIAVPSWNEQ 39
 48 MLYPRESETRVRSIDGIWNPFRSDQNPQGVDRDEWYAKELSKSRPTIMPVPASYNDI 107
 40 YOD-LCYEGRPTTYKTFPVVPPKLSQ-KHIRLYFAAVNTDCVPLNGKVGKGNHLEYLP 97
 108 TTDNLRDHVGTWVYRKFFVPRWSKQDRIMWLFSGVHYEAVWVINGQCVVHGHWGLPF 167
 98 EVDVTKVYKSG-ENELRVVVENRLKVGFP-----SKVPDSGHTVGFSGFPFPPANDFPP 152
 168 EAEVVDLISYGAENRTWMCNALLQTTVPQGRITREVPDGMIT-----VQSYTFDFPN 222
 153 YGGIIRPVLIIEFTDARILIDWDTSESEPEKKGKVKVKEVSEAVGQ-----EMTK 207
 223 YGHIRSVHLYTTPTFIEVEVTNLISK-DATVGEVFVSVNSAANEAQNVLIQIQA 281
 208 LGEEKKI---RTSNRPFGEITLBNARFW-----SLEPPYLYPLKVELEK-----DEY 253
 282 LVDKQGIIVAAVNTSQKGGKIQVNPVFWPMWYLMHSEGYLYQLEIKLATNDELDDY 341
 254 TLDIGIRITSDERKLYLNGKGVPLKGFGEKEEFPVLQGTFFPLMIKQFNLIKWNINS 313
 342 RLKVGIRITSNWSQGFPLNGKRVYFRGGRHSDIRKGLDNALMWDPFNLIKITGAN 401
 314 FRTSHYVSEEWLADRLGILVIDEADPHVGTIRYHNPETQKLAEDNIRRMIDSKHNP 373
 402 YRTSHYVSEESMOPADEHGIMWIDECPVDE--NFQGLLGGKKSLSLEQLIHRDRNP 459
 374 SVIMMSVANPEPSNHPDAEGFPKALYFTANENDRIRPVVMSMDAPPERTDVALKTYD 433

RC TISSUE=Liver;
 RX MEDLINE=99296826; PubMed=10366443; DOI=10.1006/geno.1999.5825.
 RA Fyfe J.C., Kurzhals R.L., Lasegale M.E., Henthorn P.S., Alur P.R.,
 RA Wang P., Wolte J.H., Giger U., Haskins M.E., Patterson D.F., Sun H.,
 RA Jain S., Yuhki N.;
 RT Molecular basis of feline beta-glucuronidase deficiency: an animal
 RT model of mucopolysaccharidosis VII.";
 RL Genomics 58:121-128(1999).
 CC -1- FUNCTION: Plays an important role in the degradation of dermatan
 CC and keratan sulfates.
 CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
 CC D-glucuronate.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- DISEASE: Defects in GUS are the cause of mucopolysaccharidosis
 CC type VII (MPS VII), an inherited disease reported in humans, mice,
 CC cats, and dogs.
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF012423, AAD01498.1, -;
 DR EMBL, AF012424, AAD01499.1, -;
 DR HSSP, P08236, 1BHG.
 DR Interpro, IPR008979, Gal bind like.
 DR Interpro, IPR006101, Glyco_hydro_2.
 DR Interpro, IPR006102, Glyco_hydro_2B.
 DR Interpro, IPR006104, Glyco_hydro_2B.
 DR Interpro, IPR006103, Glyco_hydro_21TM.
 DR Pfam, PF00703, Glyco_hydro_2, 1.
 DR Pfam, PF02836, Glyco_hydro_2_C, 1.
 DR Pfam, PF02837, Glyco_hydro_2_N, 1.
 DR PRINTS: PR00137; GLHYDRASE2.
 DR PROSITE, PS00719, GLYCOSYL_HYDROL_F2_1, 1.
 DR PROSITE, PS00608, GLYCOSYL_HYDROL_F2_2, 1.
 DR Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
 KW Signal.
 KW SIGNAL. 1 22 By similarity.
 FT CHAIN 23 651 Beta-glucuronidase.
 FT ACT_SITE 450 450 Proton donor (By similarity).
 FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).
 FT VARIANT 351 351 E -> K (in MPS VII).
 SO SEQUENCE 651 AA; 74609 MW; 2AE30884B70D4232 CRC64;
 Query Match 29.9%; Score 898; DB 1; Length 651;
 Best Local Similarity 36.4%; Pred. No. 1.6e-50;
 Matches 224; Conservative 98; Mismatches 212; Indels 82; Gaps 21;
 QY 4 PORNKRKRFLLINGVNLVET-SKDR-----PIAVPSWNEQYOD 42
 DB 34 PSERKE---LNGLWSPADSPSENRQGFEEQOWYTPRLRESGPTLDMVPSFPDVGCD 89
 QY 43 --LCVEEGPFTYKTFYVPKXLSQ--KHIRLYFAVAVNTDCEVFLNGEKVGKNGHLEYLPF 97
 DB 90 RQLRSFVGWVWEREATLPRQWTDGTRVLRIGSAHYALVWNGVAVAHBEGHLPF 149
 QY 98 EVDVTGKVKSG--ENELRAVVENRLKVGSPSKVP-----DSGHTVGFPGSPFPA 146
 DB 150 EADISKLVSQGPLASCRITAIINNTLT-----PHTLPPTILYQDTSKYPKYVF--VQNI 203
 QY 147 NDDFFYGGIIRPVLEFTHARILDIWDTSSSEBKKLGKVKVIEVSEAVQCEMTI 206
 DB 204 NDDFFYVAGLHRVLLVYTPPTIYIDITISTVNO-DTGLVDYQIFVEGGEHP--QLEV 259
 QY 207 KLGEERKKTIRTSNRFVEGEFLENNARFW---SLDDP-VLYPLKVEL-----KDEY 253

DB 260 RLUDSEKGVANQGTGGRQLGVNPNHLMWPYLMEHNPVLKVSLEKRLAQGRAAGSVSPFY 319
 QY 254 TLIDIGIRITSDKRLYLNKRPVLKGFKGKIEBPVLQCGFTYPLMTKDFNLKIMANS 313
 DB 320 TLPVGIRFVAATEHDFLNGKRPFYHGVNKHEDADIRSKGDFMPLVNDPMLRLWIGANA 379
 QY 314 FRTSHPYSEEMLDADLGLIVIDEAPHVGI--TRYHNPETO--TIAEDNIRMMDR 368
 DB 360 FRTSHPYAEVWQCDRYGIVVIDESPVGIVLVESYNSVLQHNLEMEELV-----R 435
 QY 369 HKNHPSVIMWVANEPEESNHPDAEGFKALYETANEMDRTPVWVWMMDAPEDETRDVA 428
 DB 436 DKNHFAVWVWVANEPAFLKPAQVYFETLHAHTALDPSRPVFTV-----NSNVEADLG 491
 QY 429 LKTFYIVCVNRYGYIYQGRIBESQLAEKDIEBLVARKKRPVTFEFGADALGIVHD 488
 DB 492 APYVDVICVNSYSGHYHDGMEVLTQLATQFEMWRYTQKPIIOSEYGDITIAGFHD 551
 QY 489 PROMSEERYOAEIVKTRILL---KDYIIGTHVWAFADPFTPONVRRPLINHGVFTR 545
 DB 552 PFLMFSESYQKGLLEQ-YHLVLDQKRYVVGELIWPADFMTNOSPQVGNKKGIFTR 610
 QY 546 DROPKLVAVHVR-RLW 560
 DB 611 QROPKGAFLRLRERYW 626
 RESULT 13
 BGLR_RAT
 ID BGLR_RAT STANDARD; PRT; 648 AA.
 AC P06760;
 DT 01-JAN-1988 (Rel. 06, Last Created)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Beta-glucuronidase precursor (EC 3.2.1.31).
 GN Name=Gusb; Synonyms=Gus;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Preputial gland;
 RX MEDLINE=87016933; PubMed=3463967;
 RA Nishimura Y., Rosenfeld M.G., Kreibich G., Gubler U., Sabatini D.D.,
 RA Adenik M., Andy R.;
 RT "Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA
 RT and in vitro insertion of its encoded polypeptide into microsomal
 RT membranes";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:7292-7296(1986).
 RN [2]
 RP SEQUENCE OF 14-648 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=88183378; PubMed=335537;
 RA Powell P.P., Kyle J.W., Miller R.D., Pantano J., Grubb J.H., Sly W.S.;
 RT "Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and
 RT expression of a chimeric protein in COS cells";
 RL Biochem. J. 250:547-555(1988).
 CC -1- FUNCTION: Plays an important role in the degradation of dermatan
 CC and keratan sulfates.
 CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
 CC D-glucuronate.
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- PTM: Undergoes a post-transcriptional proteolytic cleavage near
 CC its C-terminal end, which reduces its size by approximately 3 kDa.
 CC The site of this cleavage has as yet not been determined.
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its

SQ SEQUENCE 651 AA; 74433 MW; E8991B1E65C60120 CRC64;
 Query Match 29.6%; Score 888; DB 1; Length 651;
 Best Local Similarity 35.7%; Pred. No. 7.3e-50;
 Matches 220; Conservative 98; Mismatches 226; Indels 72; Gaps 18;
 QY 1 MVRPQNRKRFILILINGVNNLEVTSKD-----RPIAVPGSMWEO 39
 DB 27 MLYPESPSPERKDDGLMSFRADPSDGRQGEQGWYAPLRRESPTLDMVPSSFNVD 86
 QY 40 YVD--LCHEEPPYKTTTPYVPKLSQ--KHRLYFAAVNTDCEVFLNGEKYGENHLEY 94
 DB 87 GQDRLRSFVGWVWEREATLPRRWSQDGTGRVLRIGSANYAVALWVGAVHAEGGH 146
 QY 95 LPEFVDVNGKVGSG--ENELRVVVENRLKVGGFPSKV-----DSGHTTVGPPFSF 143
 DB 147 LPEFADISLVOSGLPSSCRITLAINNTLT---PPTLPPTTYKTKDSKPKGTF--V 200
 QY 144 PRANDEFPYGGIIRPVLIETDHAIRLDIWDTSSESEPEKLGKVKKIEVESEAVQGE 203
 DB 201 QNTYDFEFYVAGLHNRVLLYTTPTTYIDITVTTGVDQ-DTGLVDYQIFVQSGEHR--Q 256
 QY 204 MTKLGEERKIRTSNRVVEGEPTLENAPF---SLEDP-YLYPLKYLE-----K 250
 DB 257 LEVYLLDEBGRKVAQGTSGQRLQVFNHLMWRYLMHEHPALYLSLEVTLTAQMAAGPVS 316
 QY 251 DEYTLDIGRTISMEKRLYLNGKRPVLEKPGKHIEFPVLGGTFYPLMIKDNLKMTN 310
 DB 317 DYTTPVGRITVAVTERQFLNGKRPYFGVNGHEADIRGKGFMDPLVVKDFNLRLWG 376
 QY 311 ANSFRTSHYSEEWMLDLRLGILVIDEAPHYGI--TRYHNPETOKIAEDNIRMTDR 368
 DB 377 ANAFRTSHYSEEWMLDLRLGILVIDEAPHYGI--TRYHNPETOKIAEDNIRMTDR 368
 QY 377 ANAFRTSHYSEEWMLDLRLGILVIDEAPHYGI--TRYHNPETOKIAEDNIRMTDR 368
 DB 377 ANAFRTSHYSEEWMLDLRLGILVIDEAPHYGI--TRYHNPETOKIAEDNIRMTDR 368
 QY 369 HKNHPSVIMWSVANEPSNHPDAEGFFKALYETANEMETRPVYVMSMDADETRDYA 428
 DB 436 DKHNSVVMWSVANEPSNHPDAEGFFKALYETANEMETRPVYVMSMDADETRDYA 428
 QY 429 LKYPDVCNRYGYWYIYQGRIEEGQALEKDI BELYARHRRPIVTERGADALAGIHYD 488
 DB 492 ADYVDITCVNSYSHYHDYGHMEVIOQLATEPENWRTYQKPIIOSEGAETTINGFHD 551
 QY 489 PRPMSESEYQAELEVKTRILL--KKDYIIGTHWAFADFKTPQNVRRPILNKHGFTPR 545
 DB 552 PELMSESEYQAELEVKTRILL--KKDYIIGTHWAFADFKTPQNVRRPILNKHGFTPR 545
 QY 546 DRQPKLVANVLR-RLM 560
 DB 611 QRPKAAAFLLRRRYW 626
 RESULT 15
 Q93YV4 PRELIMINARY; PRT; 603 AA.
 AC Q93YV4; 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
 DE Beta-glucuronidase.
 GN Name=gus;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eusteroide II; Brassicales; Brassicaceae; Arabidopsie.
 OC NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C24;
 RA De Greve H., Nguyen V., Deboeck F., Thia-Toong L., Karimi M.,
 RT "T-DNA tagging of the translation initiation factor eIF-4A1 from
 Arabidopsis thaliana.";
 RL Plant Sci. 161:685-693(2001).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C24;
 RA de Greve H.,
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ298139; CAC43289.1; -
 DR EMBL; AJ298137; CAC43287.1; -
 DR HSSP; P08236; 1BHG.
 DR GO; GO:0004557; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_219.
 DR InterPro; IPR006104; Glyco_hydro_219.
 DR InterPro; IPR006103; Glyco_hydro_219.
 DR Pfam; PF00703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2; 1.
 DR Pfam; PF02837; Glyco_hydro_2; 1.
 DR PRINTS; PR00132; GLHYDRASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00609; GLYCOSYL_HYDROL_F2_2; 1.
 SQ SEQUENCE 603 AA; 68433 MW; 7562FE2EB2118DA CRC64;
 Query Match 29.4%; Score 882; DB 2; Length 603;
 Best Local Similarity 34.0%; Pred. No. 1.6e-49;
 Matches 206; Conservative 109; Mismatches 229; Indels 62; Gaps 11;
 QY 1 MVRPQNRKRFILILINGVNNLEVT-----TSKDRPIAVPGSMWEOYDLCY 45
 DB 1 MVRPQNRKRFILILINGVNNLEVT-----TSKDRPIAVPGSMWEOYDLCY 45
 QY 46 EE--GPFTYKTTFPYVPKLSQKHRLYFAAVNTDCEVFLNGEKYGENHLEYLPEFVDYTG 103
 DB 61 RNVAQNVYQREVEPIPKMAQGRIVLRPDATYHNGKAVNNQVWHEHCGTTFPADYTP 120
 QY 104 KYKSGEN-ELRVVVENRLKVGGFPSKV--DSCGHTTVGFGSPRANPDPFPYGGIIRP 159
 DB 121 YVIAKGSVRITVCNNELMQTIPPGWITDENGKKQSYF-----HDFFNVAGIHR 173
 QY 160 VLIEFTDHAIRLDIWDTSSESEPEKLGKVKKIEVESEAVQGEITILGSEBEKKIRSN 219
 DB 174 VMLYTPPTWDDITVTVHVAQ--DCNHSADVQV--VANGDVSVELRADDOQVAVTG 227
 QY 220 RFVEGEFPLENAPFWSLEDPYLPY--KYELEKDEYTLDIGRTISMEKRLYLNGKP 275
 DB 228 QGTSGTLQVNNPMLQPBEGTLYELCYAKSQTEDCDIPLAVGIRSAVAKGQFLINKP 287
 QY 276 VFLKGFGEHEEPVVGQTFYPLMIKDFNLKWINANSFRTSHYPYSEEWMLDLRLGIL 335
 DB 288 FYFTGFGHEDADLRGKGFNVLMVHDALMDWIGANSYRTSHYPAEEMLDWADDEHGIV 347
 QY 336 VIDEAPHYGI-----TRYHNPETOKIAEDNIRRMIDHKKHPSYIM 377
 DB 348 VIDETAAVGNLSLIGFGEAGNKKPELYSEAVNGETQOALQALKEILADKHPVSM 407
 QY 378 MSVANEPSNHPDAEGFFKALYETANEMETRPVYVMSMD--DAPDEKTRVALKYFDIV 435
 DB 408 WSIANEPTTRQGAAREYAPRLAEATRKDPTRPITCVNVMCDAHTOTISL----FDVL 463
 QY 436 CVNRYGYWYIYQGRIEEGQALEKDI BELYARHRRPIVTERGADALAGIHYDPOMFSE 495
 DB 464 CLNRYGYWYVOSGDELEAKYLEKEILLWQKELHQPIITTEYGVDTTLGLISMVYDMWSE 523
 QY 466 EYQAELEVKTRILLKKOYIIGTHWAFADFKTPQNVRRPILNKHGVTTRROPKLVANV 555
 DB 524 EYQCAWMLDMYHRVDSAVVAVGEOVWNPADPATSGIILRVGANKKGIPTRRDKPSAFL 583
 QY 556 LRLWMS 561
 DB 584 LQKWT 589
 Search completed: September 1, 2005, 19:32:39

Job time : 181 secs

2
.
.

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 1, 2005, 18:52:26 : Search time 28 Seconds
(without alignments) 1500.979 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001

Sequence: 1 MWRPQNRKRFLLINGVNN.....TRDROPKLVAVHRLRMSEV 563

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq1.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	4 US-09-270-957-6	Sequence 6, Appl1
2	2999	99.9	563	4 US-09-270-957-21	Sequence 21, Appl1
3	929	31.0	602	3 US-09-149-727-2	Sequence 2, Appl1
4	929	31.0	602	4 US-09-270-957-2	Sequence 2, Appl1
5	929	31.0	602	4 US-09-270-957-8	Sequence 8, Appl1
6	929	31.0	602	4 US-09-270-957-15	Sequence 15, Appl1
7	929	31.0	618	3 US-09-149-727-15	Sequence 15, Appl1
8	925	30.8	607	3 US-09-149-727-8	Sequence 8, Appl1
9	925	30.8	615	4 US-09-270-957-18	Sequence 18, Appl1
10	925	30.8	615	4 US-09-270-957-28	Sequence 28, Appl1
11	919.5	30.6	598	4 US-09-862-660-2	Sequence 2, Appl1
12	899.5	30.0	648	4 US-09-715-858-4	Sequence 12, Appl1
13	882	29.4	1010	4 US-09-118-276-12	Sequence 12, Appl1
14	879	29.3	603	3 US-09-149-727-6	Sequence 6, Appl1
15	879	29.3	603	4 US-09-270-957-17	Sequence 17, Appl1
16	879	29.3	603	4 US-09-270-957-23	Sequence 23, Appl1
17	877	29.2	832	3 US-08-630-820-7	Sequence 7, Appl1
18	877	29.2	832	4 US-09-273-453-7	Sequence 7, Appl1
19	875	29.2	604	4 US-09-893-525-37	Sequence 37, Appl1
20	875	29.2	659	4 US-09-893-525-40	Sequence 40, Appl1
21	875	29.2	850	4 US-09-893-525-42	Sequence 42, Appl1
22	872.5	29.1	602	2 US-08-882-704A-5	Sequence 5, Appl1
23	872.5	29.1	602	4 US-09-151-957-5	Sequence 5, Appl1
24	872.5	29.1	602	6 5432081-2	Patent No. 5432081
25	872.5	29.1	602	6 5432081-2	Patent No. 5432081
26	872.5	29.1	1242	4 US-09-488-270A-2	Sequence 2, Appl1
27	865.5	28.8	613	3 US-09-149-727-5	Sequence 5, Appl1

28	865.5	28.8	613	4 US-09-270-957-16	Sequence 16, Appl1
29	865.5	28.8	651	4 US-09-715-858-2	Sequence 2, Appl1
30	849.5	28.3	600	6 5268463-2	Patent No. 5268463
31	849.5	28.3	600	6 5268463-2	Patent No. 5268463
32	692.5	23.1	376	4 US-09-270-957-4	Sequence 4, Appl1
33	692.5	23.1	376	4 US-09-270-957-19	Sequence 19, Appl1
34	575	19.2	500	4 US-09-949-016-11697	Sequence 11697, A
35	574	19.1	372	4 US-09-270-957-3	Sequence 3, Appl1
36	574	19.1	372	4 US-09-270-957-22	Sequence 22, Appl1
37	496	16.5	540	4 US-09-270-957-5	Sequence 5, Appl1
38	489.5	16.3	535	4 US-09-270-957-20	Sequence 20, Appl1
39	416	13.9	1334	6 5476657-1	Patent No. 5476657
40	416	13.9	1334	6 5476657-1	Patent No. 5476657
41	414	13.8	1010	4 US-09-654-449-2	Sequence 2, Appl1
42	414	13.8	1010	4 US-09-759-152A-2	Sequence 2, Appl1
43	414	13.8	1121	1 US-07-789-915A-2	Sequence 2, Appl1
44	414	13.8	1121	1 US-08-005-002C-2	Sequence 2, Appl1
45	414	13.8	1121	1 US-08-487-203A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1					
US-09-270-957-6					
; Sequence 6, Application US/09270957					
; Patent No. 664196					
; GENERAL INFORMATION:					
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer					
; TITLE OF INVENTION: MICROBIAL, -GLUCURONIDASE GENES, GENE					
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF					
; FILE REFERENCE: 190106.405C1					
; CURRENT APPLICATION NUMBER: US/09/270,957					
; CURRENT FILING DATE: 1999-03-17					
; NUMBER OF SEQ ID NOS: 112					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 6					
; LENGTH: 563					
; TYPE: PRT					
; ORGANISM: Thermotoga maritima					
; FEATURE:					
; NAME/KEY: VARIANT					
; LOCATION: (1)-(563)					
; OTHER INFORMATION: Xaa = Any Amino Acid					
US-09-270-957-6					
Query Match					
Best local similarity 99.9%; Score 2999; DB 4; Length 563;					
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MWRPQNRKRFLLINGVNNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVVK	60		
DB	1	MWRPQNRKRFLLINGVNNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVVK	60		
QY	61	XLSQHRIRLYFAVNTDCVFLNGKVGKGNHLEVLPEVDYVKGKSGENLRVVENLT	120		
DB	61	XLSQHRIRLYFAVNTDCVFLNGKVGKGNHLEVLPEVDYVKGKSGENLRVVENLT	120		
QY	121	KVGGFSPKVPDGGTHTVGFSGFPFPPYGGIIRVVLIEFDHARILDIWDTSSG	180		
DB	121	KVGGFSPKVPDGGTHTVGFSGFPFPPYGGIIRVVLIEFDHARILDIWDTSSG	180		
QY	181	EEPKKLGKVKKIEVSEAVQEMTIKLGEEBKIRTSNRVFEGEFILENARFWSIEDPY	240		
DB	181	EEPKKLGKVKKIEVSEAVQEMTIKLGEEBKIRTSNRVFEGEFILENARFWSIEDPY	240		
QY	241	LVPKVELEKDDYTLIDIGIRITISWDEKRLYNGKVPFLKGFGBHFFVLGGTFYPLMI	300		
DB	241	LVPKVELEKDDYTLIDIGIRITISWDEKRLYNGKVPFLKGFGBHFFVLGGTFYPLMI	300		
QY	301	KQFNLLKMINANSFRTSHFYSEEWLADRLGILVIDEAPVAGITRYHYNDETOKIAD	360		
DB	301	KQFNLLKMINANSFRTSHFYSEEWLADRLGILVIDEAPVAGITRYHYNDETOKIAD	360		

QY 361 NTRMIDRKHNPVSVMVANEPSNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 DB 361 NTRMIDRKHNPVSVMVANEPSNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 QY 421 DERTDVALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 480
 DB 421 DERTDVALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 480
 QY 481 A1AGIHYPDPQMFSEEYOAEIYEKTIIRLLKKDYIIGTHVAFADPKTPQNVRRPILNKH 540
 DB 481 A1AGIHYPDPQMFSEEYOAEIYEKTIIRLLKKDYIIGTHVAFADPKTPQNVRRPILNKH 540
 QY 541 GVFTDRQPKLVAVHVRRLMSEV 563
 DB 541 GVFTDRQPKLVAVHVRRLMSEV 563

RESULT 2
 US-09-270-957-21
 ; Sequence 21, Application US/09270957
 ; Patent No. 664196
 ; GENERAL INFORMATION:
 ; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
 ; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
 ; FILE REFERENCE: 190106.405C1
 ; CURRENT APPLICATION NUMBER: US/09/270,957
 ; EARLIER FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FaastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 563
 ; TYPE: PRT
 ; ORGANISM: Thermotoga maritima
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(563)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-270-957-21

Query Match 99.9%; Score 2999; DB 4; Length 563;
 Best Local Similarity 100.0%; Pred. No. 3.8e-256;
 Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYRPOKRRKRIILINGVNNLEVTSKDRPIAVGNSWNEQYDLCYEEGPFYKTTFFVYVK 60
 DB 1 MYRPOKRRKRIILINGVNNLEVTSKDRPIAVGNSWNEQYDLCYEEGPFYKTTFFVYVK 60
 QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTKVYKSGNELRVVVENL 120
 DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTKVYKSGNELRVVVENL 120
 QY 121 KVGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 180
 DB 121 KVGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 180
 QY 121 KVGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 180
 DB 121 KVGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 180
 QY 181 BEBKLGKVKVIEVSEEAIVGEMTKLGESEKRTSNRFBEGFTILNARFWSLEDY 240
 DB 181 BEBKLGKVKVIEVSEEAIVGEMTKLGESEKRTSNRFBEGFTILNARFWSLEDY 240
 QY 241 LYPPLKVELEKDEYTDLDIGRTISWDEKRLYLNGKPVFLKFGKHEFPVLAGGTFFPLMI 300
 DB 241 LYPPLKVELEKDEYTDLDIGRTISWDEKRLYLNGKPVFLKFGKHEFPVLAGGTFFPLMI 300
 QY 301 KQFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 360
 DB 301 KQFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 360
 QY 361 NTRMIDRKHNPVSVMVANEPSNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 DB 361 NTRMIDRKHNPVSVMVANEPSNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420

QY 421 DERTDVALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 480
 DB 421 DERTDVALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 480
 QY 481 A1AGIHYPDPQMFSEEYOAEIYEKTIIRLLKKDYIIGTHVAFADPKTPQNVRRPILNKH 540
 DB 481 A1AGIHYPDPQMFSEEYOAEIYEKTIIRLLKKDYIIGTHVAFADPKTPQNVRRPILNKH 540
 QY 541 GVFTDRQPKLVAVHVRRLMSEV 563
 DB 541 GVFTDRQPKLVAVHVRRLMSEV 563

RESULT 3
 US-09-149-727-2
 ; Sequence 2, Application US/09149727
 ; Patent No. 6391547
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferson, Richard A.
 ; APPLICANT: Kilian, Andrzej
 ; APPLICANT: Keese, Paul Konrad
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
 ; FILE REFERENCE: 190106.405
 ; CURRENT APPLICATION NUMBER: US/09/149,727
 ; EARLIER FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: US 60/058,263
 ; EARLIER FILING DATE: 1997-09-09
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 602
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-149-727-2

Query Match 31.0%; Score 929; DB 3; Length 602;
 Best Local Similarity 35.6%; Pred. No. 2.5e-73;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
 QY 1 MYRPOKRRKRIILINGVNNLEVTSKDRPIAVGNSWNEQYDLCYEEGPFYKTTFFVYVK 44
 DB 1 MYRPOKRRKRIILINGVNNLEVTSKDRPIAVGNSWNEQYDLCYEEGPFYKTTFFVYVK 44
 QY 45 YEEGPFYKTTFFVYVKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTK 104
 DB 45 YEEGPFYKTTFFVYVKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTK 104
 QY 61 NHIGVYVEREFTVPAYLKQRIYLRFGSAIHKALIVNGBELVHKGGLPFEAEINNS 120
 DB 61 NHIGVYVEREFTVPAYLKQRIYLRFGSAIHKALIVNGBELVHKGGLPFEAEINNS 120
 QY 105 VKSGNELRVVVENLKVGGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 154
 DB 105 VKSGNELRVVVENLKVGGSPSKVPDSGTHVTFGFGSPFPANPFPYGGIIRPVLIETDHAHILDIWDTSS 154
 QY 121 LRDGNRRVTAVNDL-----DSTLPVGLYSRHEEGGKVRINKNPFDFNYA 170
 DB 121 LRDGNRRVTAVNDL-----DSTLPVGLYSRHEEGGKVRINKNPFDFNYA 170
 QY 155 GIIRPVLIETDHAHILDIWDTSSSEPEKLGKVKVIEVSEEAIVGEMTKLGESEKRTSNRFBEGFTILNARFWSLEDY 212
 DB 155 GIIRPVLIETDHAHILDIWDTSSSEPEKLGKVKVIEVSEEAIVGEMTKLGESEKRTSNRFBEGFTILNARFWSLEDY 212
 QY 171 GHSRPVKLYTTPFTVEDISVTDENGPF---TGVTYTVDPD---GAAETVYKSVVDBE 223
 DB 171 GHSRPVKLYTTPFTVEDISVTDENGPF---TGVTYTVDPD---GAAETVYKSVVDBE 223
 QY 213 KIRTSNRFVGEFTILNARFWSLEDYLYPLKVELEKDEYTDLDI-----GIRTSWDEK 267
 DB 213 KIRTSNRFVGEFTILNARFWSLEDYLYPLKVELEKDEYTDLDI-----GIRTSWDEK 267
 QY 224 GKVAVSTGLSGANVEIPVILMEPLNTLYIQIKVELVNDGLTIDVYEEFPGRIVEVNDG 283
 DB 224 GKVAVSTGLSGANVEIPVILMEPLNTLYIQIKVELVNDGLTIDVYEEFPGRIVEVNDG 283
 QY 268 RLYLNGKVPFLKFGKHEFPVLAGGTFFPLMIKDQFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 327
 DB 268 RLYLNGKVPFLKFGKHEFPVLAGGTFFPLMIKDQFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 327
 QY 284 KFLINNKFFYFGFKGKHDTPLNGRFGNEASVNDVFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 343
 DB 284 KFLINNKFFYFGFKGKHDTPLNGRFGNEASVNDVFNILKWINANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 343
 QY 328 LADRLGILVIDEAPVAGITRHYNDETOKIAD-----DETOKIAD-----EDNIRMRDR 368
 DB 328 LADRLGILVIDEAPVAGITRHYNDETOKIAD-----DETOKIAD-----EDNIRMRDR 368
 QY 369 HKNHPSVIMVANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 427
 DB 369 HKNHPSVIMVANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 427
 QY 401 DGNHPSVIMVANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 458
 DB 401 DGNHPSVIMVANSFRTSHYPIYSEEWLADLADRLGILVIDEAPVAGITRHYNDETOKIAD 458
 QY 428 ALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 483
 DB 428 ALKYFDIVCVNRYGYMYIYQRIIEBGLQALEKDIIEELVARRHKPIFVTEFGAD 483

```
Db      459 VAEIDVIALNRNGYVFGGDLBAKVLH---QEFHAMNKRCPGKPRIMITEYGADIVA 515
Qy      484 GIHYDPQPMFSEBYQAEVLEKTIKRLKKDYIIGTHVAFADPKTPQVNRPLNHKGVF 543
Db      516 GFHDIDPVMFTEBYQVAYQANHVVDEFENFVGQANNFADPATSQGMARVQGNKKGVF 575
Qy      544 TDRQPKLVANHRLRMSEV 563
Db      576 TDRKPKLAHVFRERWTNI 595

RESULT 4
US-09-270-957-2
; Sequence 2, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-09-270-957-2

Query Match      31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.5e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy      1 MYRPOANKRRFILLINGVNNLEV-----TSKDRPIAVPGSMN--QYQDLC 44
Db      1 MLYPINTETRGVFDLNGVNNFKLDYKGLEEKYBSKLTDTISMVPSYNDIGVTKER 60
Qy      45 YEEGPTKYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGNGHIEYLPFEVDVTGK 104
Db      61 NHIGVWYWEREFTVPAVLYKDQRIVLRFSGATHKAIYVNGELVHEHKGFLFPEAEINNS 120
Qy      105 VSGENELRVVVENRLKVGFPSPKVPDSGTHVGFSG-----PPANPDPFPYG 154
Db      121 LADGNMRVTVAVDNL-----DSTLPVGLYSERHEBGLGVIRNKNPFDFENYA 170
Qy      155 GIIRPVLEFTHARLIDIVDTSSEPEPKLGGKVKVKEVSEBAVGEMTIKLG--EEE 212
Db      171 GHRPVKITYTPTFYVEDISVTDENG--TGTVTYVDFO---GKAETVKVSVVDEE 223
Qy      213 KIRTSNRFVEGEFILLENAFWSLEDPLYPLKVELKEDEYTLDI-----GIRTSWDEK 267
Db      224 GKVAVSTEGLSGNVEIPNVILMEPLNTIYQIKVELVNDGLTIIDYEEPFGRIVAVNDG 283
Qy      268 RLYLNGKPVFLKGFQKHEEPVVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEMLD 327
Db      284 KFLINNKPFYFGFGKHEDTPIINGRGFNBSVMVDNFIKMGANSFRTAHAPYSEELMR 343
Qy      328 LADRLGILVIDAPRHVIGTRHYN-----PETQKIA-----ENIRMRIDR 368
Db      344 LADRBGLVVIDETPAVG--HLMFMATTGLGEGSERVSTWEEKIRTFEHHQVLELVSF 400
Qy      369 HKNHPSVIMSVANBESNHPDAEGFFKALYETANEMD--RTRPVVWVSMMDAPDERTRDV 427
Db      401 DKNHPSVVMWSIANBAATEEBGAYEFKPLVELTELDPQKRPTIIVLFVMAVTP--TDK 458
Qy      428 ALKYDPDIVCVNRYYGYIYQGRIBGLQALEKDIEELVARNH---KPIFVTEFGADALA 483
Db      459 VAEIDVIALNRNGYVFGGDLBAKVLH---QEFHAMNKRCPGKPRIMITEYGADIVA 515
Qy      484 GIHYDPQPMFSEBYQAEVLEKTIKRLKKDYIIGTHVAFADPKTPQVNRPLNHKGVF 543
Db      576 TDRKPKLAHVFRERWTNI 595
```

```
Db      516 GFHDIDPVMFTEBYQVAYQANHVVDEFENFVGQANNFADPATSQGMARVQGNKKGVF 575
Qy      544 TDRQPKLVANHRLRMSEV 563
Db      576 TDRKPKLAHVFRERWTNI 595

RESULT 5
US-09-270-957-8
; Sequence 8, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 602
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-09-270-957-8

Query Match      31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.5e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy      1 MYRPOANKRRFILLINGVNNLEV-----TSKDRPIAVPGSMN--QYQDLC 44
Db      1 MLYPINTETRGVFDLNGVNNFKLDYKGLEEKYBSKLTDTISMVPSYNDIGVTKER 60
Qy      45 YEEGPTKYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGNGHIEYLPFEVDVTGK 104
Db      61 NHIGVWYWEREFTVPAVLYKDQRIVLRFSGATHKAIYVNGELVHEHKGFLFPEAEINNS 120
Qy      105 VSGENELRVVVENRLKVGFPSPKVPDSGTHVGFSG-----PPANPDPFPYG 154
Db      121 LADGNMRVTVAVDNL-----DSTLPVGLYSERHEBGLGVIRNKNPFDFENYA 170
Qy      155 GIIRPVLEFTHARLIDIVDTSSEPEPKLGGKVKVKEVSEBAVGEMTIKLG--EEE 212
Db      171 GHRPVKITYTPTFYVEDISVTDENG--TGTVTYVDFO---GKAETVKVSVVDEE 223
Qy      213 KIRTSNRFVEGEFILLENAFWSLEDPLYPLKVELKEDEYTLDI-----GIRTSWDEK 267
Db      224 GKVAVSTEGLSGNVEIPNVILMEPLNTIYQIKVELVNDGLTIIDYEEPFGRIVAVNDG 283
Qy      268 RLYLNGKPVFLKGFQKHEEPVVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEMLD 327
Db      284 KFLINNKPFYFGFGKHEDTPIINGRGFNBSVMVDNFIKMGANSFRTAHAPYSEELMR 343
Qy      328 LADRLGILVIDAPRHVIGTRHYN-----PETQKIA-----ENIRMRIDR 368
Db      344 LADRBGLVVIDETPAVG--HLMFMATTGLGEGSERVSTWEEKIRTFEHHQVLELVSF 400
Qy      369 HKNHPSVIMSVANBESNHPDAEGFFKALYETANEMD--RTRPVVWVSMMDAPDERTRDV 427
Db      401 DKNHPSVVMWSIANBAATEEBGAYEFKPLVELTELDPQKRPTIIVLFVMAVTP--TDK 458
Qy      428 ALKYDPDIVCVNRYYGYIYQGRIBGLQALEKDIEELVARNH---KPIFVTEFGADALA 483
Db      459 VAEIDVIALNRNGYVFGGDLBAKVLH---QEFHAMNKRCPGKPRIMITEYGADIVA 515
Qy      516 GFHDIDPVMFTEBYQVAYQANHVVDEFENFVGQANNFADPATSQGMARVQGNKKGVF 575
Db      544 TDRQPKLVANHRLRMSEV 563
Db      576 TDRKPKLAHVFRERWTNI 595
```


; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 28
 ; LENGTH: 615
 ; TYPE: PRF
 ; ORGANISM: *Bacillus* sp.
 US-09-270-957-28

Query Match 30.8%; Score 925; DB 4; Length 615;
 Best Local Similarity 35.7%; Pred. No. 5.9e-73;
 Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

QY 4 PQRNKKRFFLLINGVNLLEVT-----TSKDRP-----IAPGSMNE--QYODLCYEE 47
DB 9 PINTREKGVFDLNGVNLNFKLDYDKGLSEKWTYSKLTDTTSMNVPSSYNDIGYTKKIRNHI 68
QY 48 GPFYKTFYVPKLSQKHILRYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKYS 107
DB 69 GYVWVEREFTVPAYLKQRIVLRFSGATHKAIYVVGELVVEHKGGLPFEALINSLRD 128
QY 108 GENELRVVENRLKVGSPSKVPDSGTHVGFSG-----PPRANPFPFYGGIT 157
DB 129 GGNRVYVAVDNLT-----DSTLPVGLYSERHEGLKVIRNKNPDPFFNYAGIH 178
QY 158 RPLVEFTDHAFLIDIVDTSSEPEKKLGKVKIEVSEBAVGQEMTKLG--EEEKKI 215
DB 179 RYVKYITTFYTVEDISVTTDNGP--TGIVTYTVDFQ---GKAETVKSVDVEBGKV 231
QY 216 RTSNRFVEGEFLLENARFWSLEDPLYLPLKVELKDEYTLDI-----GIRTSMDERLY 270
DB 222 VASTGLSGNVEIPNVLWEPLNTYLYQKVELVNDGLTIYEEBPGVRYTVNDGKFL 291
QY 271 LKNGKVEFLKFGKHEFPVLGCGTFYPLMKDNLKWINANSFRTSHYPSSEWLDLAD 330
DB 292 INNKPFYFGFGKHEDTPIKNGRGNFNASVMDENILKMGANSFPAHAYSEELMRAD 351
QY 331 RLGILVIDAPHYGATRYHYN-----PETOKIA-----EENIRMI DRHN 371
DB 352 RGLGVVIDTPAVGV---HINFMATTGLGSESERSTWETKIRTFEHHQDVLELVSBDKN 408
QY 372 HBSVIMSVANBESNHPDAEGFKALEYETANEMD--RTRPVVMVSMMDAPDERTRDVALK 430
DB 409 HBSVMMSTIANBAATEBEGAYEFKPLVELTKELDPQKRPVTVILFVMAVTP--TDKVAE 466
QY 431 YFDIVCVARYGMYIYQGIIEEGQALEDDIELVARNH---KPLFTEBADAALAGH 486
DB 467 LIDVIALNRYNGWYFDGDLAQAHLR---QEFHAMNRCRCPKPIMTIEYGADTVAGFH 523
QY 487 YDPQWSEBYOAEIVKTRILLLKKDYIIGTHVAFADFTTPONVRRPILNHHGVFTPD 546
DB 524 DIDPMFTEYQVEYQAHNVVDFDEFENVGEOANFADPATSQGYMRVQGGKGVFTRD 583
QY 547 RQPKLVAVHLRLMSEV 563
DB 584 RKPKLAAHVFRERMTWI 600
  
```

RESULT 11
 US-09-862-660-2
 ; Sequence 2, Application US/09862660
 ; Patent No. 6664097
 ; GENERAL INFORMATION:
 ; APPLICANT: Russell, William
 ; APPLICANT: Klaenhammer, Todd
 ; TITLE OF INVENTION: LACTOBACILLUS BETA-GLUCURONIDASE AND DNA ENCODING THE SAME
 ; FILE REFERENCE: 5051.514
 ; CURRENT APPLICATION NUMBER: US/09/862,660
 ; CURRENT FILING DATE: 2001-05-21
 ; PRIOR APPLICATION NUMBER: 60/206,372
 ; PRIOR FILING DATE: 2000-05-23
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2
 ; LENGTH: 598
 ; TYPE: PRF
 ; ORGANISM: *Lactobacillus gaseeri*
 US-09-862-660-2

Query Match 30.6%; Score 919.5; DB 4; Length 598;
 Best Local Similarity 34.1%; Pred. No. 1.7e-72;
 Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;

```

QY 4 PQRNKKRFFLLINGVNLLEVT-----SKDRP-----IAPGSMNE--QYODLCYEE 47
DB 7 PIONKRYRENTLMNGWQFETDPNSVGLDEGNKKELPDEEMVPQGTFAELTTKDRKYYT 66
QY 48 GPFYKTFYVPKLSQKHILRYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKYS 107
DB 67 GDFWYQKDFEFLPSFLKKELIYRFGSVYHRAKVPFLNGHVQHGGLPFGYKSNLYNY 126
QY 108 GE-NELRVVENRLKVGSP--SKVPDSGTHVGFSGPPRANPFPFYGGIIRPVLE 163
DB 127 DQTRVYVVLVNNLESEKAIPEGTEILDNGQ-----KLAQPYDFPNYSGIMANVLL 179
QY 164 FTDHARLIDIVDTSSEPEKKLGKVKIEVSEBAVGQ--EMTKLGSEKKIRTSNRFV 222
DB 180 ALPOSQITNFKLN-----YQANNKATITYNIEANNNAEFKVTLPDNOKEVACATSKN 232
QY 223 EGEFLLENARFWSLEDPLYLPLKYLEK-----DEYTLDIGIRTSMDERLYLNGKRVF 277
DB 233 TSSLITIKNPLMSPDPYSYKIKIEMLEDGTVBEYTDKIGIRTVKIVNDKILLNNHPIY 292
QY 278 LKFGKHEFPVLGCGTFYPLMKDNLKWINANSFRTSHYPSSEWLDLADRLGILVI 337
DB 292 LKFGKHEFPVLGKAVNESIIRKDYECMKKIGANCFSSHYPFAEBEYQVADKGFLLI 352
QY 338 DEAPHVGITRYHYN-----PETOKIAEDNIRMI DRHNHPSVIMWS 379
DB 353 DEVPAGVGNRSITFPLVNTNSQGHFASKTVPELKKVHEQOIKEMIRDRHPSEVIAMS 412
QY 380 VANBESNHPDAEGFKALEYETANEMD--RTRPVVMVSMMDAPDERTRDVALKVPDYVN 438
DB 413 LFNPESTQSSYDFKDIFAFKLDPQNPBYGTLMWSPK--VKKLHPLCDFVCLN 470
QY 439 RYVGYIYQG--RIEGLQALEKDIIELY--ARRKPIFVTEGADALAGIHYDPQWSE 496
DB 471 RYVGYVAGGPEIYNNAKMWLEDELQGNLKNRPVTEFGADTLSSSHLLPDEMSQE 530
QY 497 YQAEIVKTRILLLKKDYIIGTHVAFADFTTPONVRRPILNHHGVFTPD 556
DB 531 YQNEYYQWYFDIFKKYPPICGELVWNFADFTSSEGIMRGVGNQDGIPTDRREPDIAPFL 590
QY 557 RLHSEV 563
DB 591 KKRWOQL 597
  
```

RESULT 12
 US-09-715-858-4
 ; Sequence 4, Application US/09715858
 ; Patent No. 6582692
 ; GENERAL INFORMATION:
 ; APPLICANT: Podaskoff, Gregory
 ; APPLICANT: Watson, Gordon
 ; APPLICANT: Couto, Linda B.
 ; APPLICANT: Yang, Bin
 ; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
 ; TREATMENT OF LYSSOMAL DISORDERS
 ; FILE REFERENCE: 0800-0021
 ; CURRENT APPLICATION NUMBER: US/09/715,858
 ; CURRENT FILING DATE: 2000-11-14
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 648


```
US-09-149-727-6
; Sequence 6, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 603
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-149-727-6
```

Query Match 29.3%; Score 879; DB 3; Length 603;

Best Local Similarity 33.8%; Pred. No. 6.5e-69;

Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

```
QY 1 MVRPQNKRRFILLINGVNNLEV-----TSKDRPIAVGSMNEQYODLCY 45
DB 1 MVRPQNKRRFILLINGVNNLEV-----TSKDRPIAVGSMNEQYODLCY 60
QY 46 EE--GPFTYKTFYVVKLSQKHIRLYFAAVNTDCEVFLNGKEGVENHIEYLPEFVDVYG 103
DB 61 RNYAGNWWYQREVFIPKMGAGQRIYRFDVATTHYGKVVNNQEWHEHGGYTFPEADVTP 120
QY 104 KYKSGEN-ELRVVENRNLKVGPPSKV--DPSGTHTVGFPGSPPPANFDFPPYGGIIRP 159
DB 121 VYIAGKSVRIYVCNNELNMQTIIPGMVITDENGKKQSYF-----HDFPNVAGIHRS 173
QY 160 VLIEFTDARILDIWDTSESEPEKKLGKVKYKIEVSEAVGOEMTILGSEBEKKIRISN 219
DB 174 VMLYTPPTWDDITVYVTHAQ--DCNHSVDMQV--VANGDVSVELRADQOVVATG 227
QY 220 RFVEGEFLEENARFMSLEDPYLPYPL---KVELEKDEYTLDIGRTISWDEKRLYLNKP 275
DB 228 QGTSGTLQVNPMLMQPGEGLYELCVTAKSQTECDIYPLRVGIRSAVAKGEOFLINKP 287
QY 276 VFLKGFGRKEEFPVVGQTFYPLMIKDENVLKWINANSFRTSHYPSSEMLDLADRLGIL 335
DB 288 FYFTGFGRHEDADLRGKGFNDVLMVHDALMDWIGANSYRSHYPAEEMLDMADEHGIV 347
QY 336 VIDEAPHVGI-----TRYHNPETOKIADNIRRMIDRHKHPSYIM 377
DB 348 VIDETAAVGFNLISLIGFEAGNKKPELYSEEAUVGETOQAHLOAIKELIARDKHNPSYVM 407
QY 378 MSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTVALKYPDI 435
DB 408 WSIANEPTDRPGAREYFAPLAETARKLDPTRPITCVNMVPCDAHTDITISD-----FDVL 463
QY 436 CVNRYYGYWYIYQRIEGLQALEKDIIEELVANHRRKPIFVTEFGADALAGIHYDPOMSE 495
DB 464 CLNRYGYWYVQSGDLETAKEVLEKELIWAQEKLHQPIITTEYGVDTLAGLSMYTDMSE 523
QY 496 EYQAEIVKTRILLKDDYIIGTHVMAFADFTKQVNRPRPILNKHGVFTRDQPLVAVH 555
DB 524 EYQCAMLDMYHRVFDKSAVAVGQVWNPADPATSQILRVGNGKKGIPTDRKPKSAFL 583
QY 556 LRLRWS 561
DB 584 LQKRW 589
```

RESULT 15

US-09-270-957-17

```
; Sequence 17, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; APPLICANT: MICROSIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 603
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-270-957-17
```

Query Match 29.3%; Score 879; DB 4; Length 603;

Best Local Similarity 33.8%; Pred. No. 6.5e-69;

Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

```
QY 1 MVRPQNKRRFILLINGVNNLEV-----TSKDRPIAVGSMNEQYODLCY 45
DB 1 MVRPQNKRRFILLINGVNNLEV-----TSKDRPIAVGSMNEQYODLCY 60
QY 46 EE--GPFTYKTFYVVKLSQKHIRLYFAAVNTDCEVFLNGKEGVENHIEYLPEFVDVYG 103
DB 61 RNYAGNWWYQREVFIPKMGAGQRIYRFDVATTHYGKVVNNQEWHEHGGYTFPEADVTP 120
QY 104 KYKSGEN-ELRVVENRNLKVGPPSKV--DPSGTHTVGFPGSPPPANFDFPPYGGIIRP 159
DB 121 VYIAGKSVRIYVCNNELNMQTIIPGMVITDENGKKQSYF-----HDFPNVAGIHRS 173
QY 160 VLIEFTDARILDIWDTSESEPEKKLGKVKYKIEVSEAVGOEMTILGSEBEKKIRISN 219
DB 174 VMLYTPPTWDDITVYVTHAQ--DCNHSVDMQV--VANGDVSVELRADQOVVATG 227
QY 220 RFVEGEFLEENARFMSLEDPYLPYPL---KVELEKDEYTLDIGRTISWDEKRLYLNKP 275
DB 228 QGTSGTLQVNPMLMQPGEGLYELCVTAKSQTECDIYPLRVGIRSAVAKGEOFLINKP 287
QY 276 VFLKGFGRKEEFPVVGQTFYPLMIKDENVLKWINANSFRTSHYPSSEMLDLADRLGIL 335
DB 288 FYFTGFGRHEDADLRGKGFNDVLMVHDALMDWIGANSYRSHYPAEEMLDMADEHGIV 347
QY 336 VIDEAPHVGI-----TRYHNPETOKIADNIRRMIDRHKHPSYIM 377
DB 348 VIDETAAVGFNLISLIGFEAGNKKPELYSEEAUVGETOQAHLOAIKELIARDKHNPSYVM 407
QY 378 MSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTVALKYPDI 435
DB 408 WSIANEPTDRPGAREYFAPLAETARKLDPTRPITCVNMVPCDAHTDITISD-----FDVL 463
QY 436 CVNRYYGYWYIYQRIEGLQALEKDIIEELVANHRRKPIFVTEFGADALAGIHYDPOMSE 495
DB 464 CLNRYGYWYVQSGDLETAKEVLEKELIWAQEKLHQPIITTEYGVDTLAGLSMYTDMSE 523
QY 496 EYQAEIVKTRILLKDDYIIGTHVMAFADFTKQVNRPRPILNKHGVFTRDQPLVAVH 555
DB 524 EYQCAMLDMYHRVFDKSAVAVGQVWNPADPATSQILRVGNGKKGIPTDRKPKSAFL 583
QY 556 LRLRWS 561
DB 584 LQKRW 589
```

Search completed: September 1, 2005, 19:14:16

Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:04:53 ; Search time 167 Seconds

(without alignments)
1327.679 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MVRPQNKRRFILLINGVWN.....TRDROPKLVAYHARIMSEV 563Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1774312 seqs, 393823214 residues

Total number of hits satisfying chosen parameters: 1774312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCRT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/PCRTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US10F_PUBCOMB.pep.*
19: /cgn2_6/prodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2999	99.9	563	US-10-364-649-6	Sequence 6, Appl1
2	2999	99.9	563	US-10-364-649-21	Sequence 21, Appl1
3	2999	99.9	563	US-10-369-493-2995	Sequence 2995, Ap
4	2999	99.9	563	US-10-757-093-20	Sequence 20, Appl1
5	1011	33.7	570	US-10-757-093-19	Sequence 19, Appl1
6	929	31.0	602	US-10-120-145-2	Sequence 2, Appl1
7	929	31.0	602	US-10-364-649-2	Sequence 2, Appl1
8	929	31.0	602	US-10-364-649-8	Sequence 8, Appl1
9	929	31.0	602	US-10-364-649-15	Sequence 15, Appl1
10	929	31.0	602	US-10-757-093-23	Sequence 23, Appl1
11	929	31.0	618	US-10-120-145-4	Sequence 4, Appl1

12	925	30.8	607	US-10-120-145-8	Sequence 8, Appl1
13	925	30.8	607	US-10-364-649-18	Sequence 18, Appl1
14	925	30.8	615	US-10-364-649-28	Sequence 28, Appl1
15	920.5	30.7	641	US-10-757-093-2	Sequence 2, Appl1
16	919.5	30.6	598	US-09-862-660-2	Sequence 2, Appl1
17	919.5	30.6	598	US-10-673-935-2	Sequence 2, Appl1
18	919.5	30.6	598	US-10-757-093-21	Sequence 21, Appl1
19	905	30.2	670	US-10-757-093-12	Sequence 12, Appl1
20	902.5	30.1	648	US-10-757-093-13	Sequence 13, Appl1
21	899.5	30.0	641	US-10-421-175-4	Sequence 4, Appl1
22	898	29.9	658	US-10-757-093-15	Sequence 15, Appl1
23	894.5	29.8	648	US-10-757-093-14	Sequence 14, Appl1
24	888	29.6	651	US-10-757-093-16	Sequence 16, Appl1
25	883	29.4	634	US-10-757-093-4	Sequence 4, Appl1
26	882	29.4	1010	US-09-118-276-12	Sequence 12, Appl1
27	882	29.4	1010	US-10-705-197A-12	Sequence 12, Appl1
28	879	29.3	603	US-10-161-403-106	Sequence 106, App
29	879	29.3	603	US-10-120-145-6	Sequence 6, Appl1
30	879	29.3	603	US-10-364-649-17	Sequence 17, Appl1
31	879	29.3	603	US-10-364-649-23	Sequence 23, Appl1
32	879	29.3	603	US-10-369-493-842	Sequence 842, App
33	879	29.3	603	US-10-161-408-17	Sequence 17, Appl1
34	879	29.3	603	US-10-757-093-22	Sequence 22, Appl1
35	879	29.3	603	US-11-006-076-106	Sequence 106, App
36	877	29.2	618	US-10-356-088-28	Sequence 28, Appl1
37	877	29.2	618	US-10-799-326-28	Sequence 28, Appl1
38	877	29.2	832	US-10-632-815-7	Sequence 7, Appl1
39	875	29.2	604	US-09-893-525-37	Sequence 37, Appl1
40	875	29.2	604	US-10-763-380-37	Sequence 37, Appl1
41	875	29.2	659	US-09-893-525-40	Sequence 40, Appl1
42	875	29.2	659	US-10-763-380-40	Sequence 40, Appl1
43	875	29.2	850	US-09-893-525-42	Sequence 42, Appl1
44	875	29.2	850	US-10-763-380-42	Sequence 42, Appl1
45	872.5	29.1	602	US-10-195-518-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-10-364-649-6
; Sequence 6, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106 405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-6

Query Match 99.9%; Score 2999; DB 15; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.9e-241;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVRPQNKRRFILLINGVWNLEVTSKDRPIAYPGSMNEGYODLCYEEBGFYTKTFFVYVK 60
Db 1 MVRPQNKRRFILLINGVWNLEVTSKDRPIAYPGSMNEGYODLCYEEBGFYTKTFFVYVK 60

QY 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120
 DB 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120
 QY 121 KVGFPSPKVPDGGTHVGFSGFPANPDPFPYGGIIRPVLIIEPTDHAIRLIDIVWDTSES 180
 DB 121 KVGFPSPKVPDGGTHVGFSGFPANPDPFPYGGIIRPVLIIEPTDHAIRLIDIVWDTSES 180
 QY 181 EPEKLGKVKKIVSEEAQVQEMTIKLGEEBKIRTSNRPEGEFIIENARFWSLEDPY 240
 DB 181 EPEKLGKVKKIVSEEAQVQEMTIKLGEEBKIRTSNRPEGEFIIENARFWSLEDPY 240
 QY 241 LYPKLEKEDEYTLIDIGIRITISWDEKRLYNGKRVFLKGGKEFPVLGGGTFFPLMI 300
 DB 241 LYPKLEKEDEYTLIDIGIRITISWDEKRLYNGKRVFLKGGKEFPVLGGGTFFPLMI 300
 QY 301 KDFNLKMINANSFRTSHYPYSEEWLADRLGLIVIDEAPHVGITRHYNPETQKIAD 360
 DB 301 KDFNLKMINANSFRTSHYPYSEEWLADRLGLIVIDEAPHVGITRHYNPETQKIAD 360
 QY 361 NRRMIDRKHNPVSIVMSVANEPESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 DB 361 NRRMIDRKHNPVSIVMSVANEPESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 QY 421 DERTDVALKYFDIVCVNRYGMYIYQGRIEEGLOALEKDIIBELVARRKPIFVTEFGAD 480
 DB 421 DERTDVALKYFDIVCVNRYGMYIYQGRIEEGLOALEKDIIBELVARRKPIFVTEFGAD 480
 QY 481 AIAGIHYDPQPMFSEBYQAEIYEKTIIRLLKKDYIIGTHVMAFADFKTPQVNRRLINMK 540
 DB 481 AIAGIHYDPQPMFSEBYQAEIYEKTIIRLLKKDYIIGTHVMAFADFKTPQVNRRLINMK 540
 QY 541 GVFTDRQPKLVAVHRLRLMSEV 563
 DB 541 GVFTDRQPKLVAVHRLRLMSEV 563

RESULT 2
 US-10-364-649-21
 ; Sequence 21, Application US/10364649
 ; Publication No. US20030229921A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Richard A. Jefferson and Jorge B. Mayer
 ; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
 ; FILE REFERENCE: 190106.405C1
 ; CURRENT APPLICATION NUMBER: US/10/364,649
 ; CURRENT FILING DATE: 2003-02-12
 ; PRIOR APPLICATION NUMBER: 10/364,649
 ; PRIOR FILING DATE: 2003-02-12
 ; PRIOR APPLICATION NUMBER: US 09/270,957
 ; PRIOR FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 563
 ; TYPE: PRT
 ; ORGANISM: Thermotoga maritima
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(563)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-10-364-649-21

Query Match 99.9%; Score 2999; DB 15; Length 563;
 Best Local Similarity 100.0%; Pred. No. 1.9e-241;
 Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRPQNRKGRFILLINGVNLVETSKDRPIAVPGSMNEQYODLCYEBGPFYKTTFFVYPK 60
 DB 1 MVRPQNRKGRFILLINGVNLVETSKDRPIAVPGSMNEQYODLCYEBGPFYKTTFFVYPK 60

QY 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120
 DB 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120
 QY 121 KVGFPSPKVPDGGTHVGFSGFPANPDPFPYGGIIRPVLIIEPTDHAIRLIDIVWDTSES 180
 DB 121 KVGFPSPKVPDGGTHVGFSGFPANPDPFPYGGIIRPVLIIEPTDHAIRLIDIVWDTSES 180
 QY 181 EPEKLGKVKKIVSEEAQVQEMTIKLGEEBKIRTSNRPEGEFIIENARFWSLEDPY 240
 DB 181 EPEKLGKVKKIVSEEAQVQEMTIKLGEEBKIRTSNRPEGEFIIENARFWSLEDPY 240
 QY 241 LYPKLEKEDEYTLIDIGIRITISWDEKRLYNGKRVFLKGGKEFPVLGGGTFFPLMI 300
 DB 241 LYPKLEKEDEYTLIDIGIRITISWDEKRLYNGKRVFLKGGKEFPVLGGGTFFPLMI 300
 QY 301 KDFNLKMINANSFRTSHYPYSEEWLADRLGLIVIDEAPHVGITRHYNPETQKIAD 360
 DB 301 KDFNLKMINANSFRTSHYPYSEEWLADRLGLIVIDEAPHVGITRHYNPETQKIAD 360
 QY 361 NRRMIDRKHNPVSIVMSVANEPESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 DB 361 NRRMIDRKHNPVSIVMSVANEPESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
 QY 421 DERTDVALKYFDIVCVNRYGMYIYQGRIEEGLOALEKDIIBELVARRKPIFVTEFGAD 480
 DB 421 DERTDVALKYFDIVCVNRYGMYIYQGRIEEGLOALEKDIIBELVARRKPIFVTEFGAD 480
 QY 481 AIAGIHYDPQPMFSEBYQAEIYEKTIIRLLKKDYIIGTHVMAFADFKTPQVNRRLINMK 540
 DB 481 AIAGIHYDPQPMFSEBYQAEIYEKTIIRLLKKDYIIGTHVMAFADFKTPQVNRRLINMK 540
 QY 541 GVFTDRQPKLVAVHRLRLMSEV 563
 DB 541 GVFTDRQPKLVAVHRLRLMSEV 563

RESULT 3
 US-10-369-493-2995
 ; Sequence 2995, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 2995
 ; LENGTH: 563
 ; TYPE: PRT
 ; ORGANISM: Thermotoga maritima
 US-10-369-493-2995

Query Match 99.9%; Score 2999; DB 15; Length 563;
 Best Local Similarity 99.8%; Pred. No. 1.9e-241;
 Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVRPQNRKGRFILLINGVNLVETSKDRPIAVPGSMNEQYODLCYEBGPFYKTTFFVYPK 60
 DB 1 MVRPQNRKGRFILLINGVNLVETSKDRPIAVPGSMNEQYODLCYEBGPFYKTTFFVYPK 60

QY 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120
 DB 61 XLSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPEFVDYTGKVKSGENELRVVVENRL 120

```

Qy 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 180
Db 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 180
Qy 181 BEBKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 240
Db 181 BEBKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 240
Qy 241 LVPKLELEKDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEFPVLGGTFYPLMI 300
Db 241 LVPKLELEKDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEFPVLGGTFYPLMI 300
Qy 301 KOFNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAPHYGTRRYANPETOKIAD 360
Db 301 KOFNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAPHYGTRRYANPETOKIAD 360
Qy 361 NIRMIDRKHNPSTVIMSVANESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
Db 361 NIRMIDRKHNPSTVIMSVANESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
Qy 421 DERTDVALKYPDIYCVNRYGMYIYQGRIBEGLOALEKODIEBLYARHRKPIFVTEFGAD 480
Db 421 DERTDVALKYPDIYCVNRYGMYIYQGRIBEGLOALEKODIEBLYARHRKPIFVTEFGAD 480
Qy 481 A1AGIHYPDPOMFSEBYQALVEKTIIRLLKKOYIIGTHVMAFADPKTPQNRRPILNKH 540
Db 481 A1AGIHYPDPOMFSEBYQALVEKTIIRLLKKOYIIGTHVMAFADPKTPQNRRPILNKH 540
Qy 541 GVFTDRQPKLVAAHVLRLMSEV 563
Db 541 GVFTDRQPKLVAAHVLRLMSEV 563

```

```

RESULT 4
US-10-757-093-20
; Sequence 20, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757,093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent version 3.2
; SEQ ID NO 20
; LENGTH: 563
; TYPE: PRF
; ORGANISM: Thermotoga maritima
US-10-757-093-20

```

```

Query Match 99.9%; Score 2999; DB 18; Length 563;
Best Local Similarity 99.8%; Pred. No. 1.9e-241;
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRPQANKRFFILLINGVNNLEVTSDRPIAVPGSWNEQYODLCYEEGFFTYKTFYVVK 60
Db 1 MRPQANKRFFILLINGVNNLEVTSDRPIAVPGSWNEQYODLCYEEGFFTYKTFYVVK 60
Qy 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGBNHIETLPEFVDVTGKYSGENELRVVVENL 120
Db 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGBNHIETLPEFVDVTGKYSGENELRVVVENL 120
Qy 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 180
Db 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 180
Qy 181 BEBKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 240
Db 181 BEBKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 240
Qy 241 LVPKLELEKDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEFPVLGGTFYPLMI 300
Db 241 LVPKLELEKDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEFPVLGGTFYPLMI 300

```

```

Db 241 LVPKLELEKDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEFPVLGGTFYPLMI 300
Qy 301 KOFNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAPHYGTRRYANPETOKIAD 360
Db 301 KOFNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAPHYGTRRYANPETOKIAD 360
Qy 361 NIRMIDRKHNPSTVIMSVANESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
Db 361 NIRMIDRKHNPSTVIMSVANESNHPDAGFFKALYETANEMDRTPVVMVSMMDAP 420
Qy 421 DERTDVALKYPDIYCVNRYGMYIYQGRIBEGLOALEKODIEBLYARHRKPIFVTEFGAD 480
Db 421 DERTDVALKYPDIYCVNRYGMYIYQGRIBEGLOALEKODIEBLYARHRKPIFVTEFGAD 480
Qy 481 A1AGIHYPDPOMFSEBYQALVEKTIIRLLKKOYIIGTHVMAFADPKTPQNRRPILNKH 540
Db 481 A1AGIHYPDPOMFSEBYQALVEKTIIRLLKKOYIIGTHVMAFADPKTPQNRRPILNKH 540
Qy 541 GVFTDRQPKLVAAHVLRLMSEV 563
Db 541 GVFTDRQPKLVAAHVLRLMSEV 563

```

```

RESULT 5
US-10-757-093-19
; Sequence 19, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757,093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent version 3.2
; SEQ ID NO 19
; LENGTH: 570
; TYPE: PRF
; ORGANISM: Sulfolobus solfataricus
US-10-757-093-19

```

```

Query Match 33.7%; Score 1011; DB 18; Length 570;
Best Local Similarity 39.2%; Pred. No. 2.2e-75;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

Qy 15 LINGVNNLEVTSDRPIAVPGSWNEQYODLCYEEGFFTYKTFYVVK 60
Db 15 LINGVNNLEVTSDRPIAVPGSWNEQYODLCYEEGFFTYKTFYVVK 60
Qy 11 LGGFWKFKINDMENTGEBNGWYKLGSESDIIVPASMNQNPQMDQFSGIAYQKDLFVSN 70
Db 11 LGGFWKFKINDMENTGEBNGWYKLGSESDIIVPASMNQNPQMDQFSGIAYQKDLFVSN 70
Qy 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGBNHIETLPEFVDVTGKYSGENELRVVVENL 120
Db 61 XLSQKHRLYFAAVNTDCEVFLNGEKVGBNHIETLPEFVDVTGKYSGENELRVVVENL 120
Qy 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 174
Db 121 KVGSPSKVPDSGTHVGFSGFPANPDPFYGGIIRPVLEFTDHARILDIWDTSS 174
Qy 174 VDTSEBEKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 230
Db 174 VDTSEBEKLGKVKVKEVSEAVGDMTIKLGEEBKIRTSNRPVEGEFPLENARFWSLEDY 230
Qy 231 ARFWSLEDYPLKVELE---KDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEH 285
Db 231 ARFWSLEDYPLKVELE---KDEYTLDIGRTISWDEKRLYLNGKRPVFLKGFGEH 285
Qy 286 BPPVVGOGFPYPLMKDNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAP--HV 343
Db 286 BPPVVGOGFPYPLMKDNLKWINANSFRTSHPYSEEWDLADRGLIIVDEAP--HV 343
Qy 344 GTRRYANPETOKI-----AEDNTRMIDRKHNPSTVIMSVANESNHPDAGGF 394
Db 344 GTRRYANPETOKI-----AEDNTRMIDRKHNPSTVIMSVANESNHPDAGGF 394
Qy 394 NISRVMSGEALAKMGDVKYFEKAVDTIKEMIRQHKNPSTVIMSVANESNHPDAGGF 402
Db 394 NISRVMSGEALAKMGDVKYFEKAVDTIKEMIRQHKNPSTVIMSVANESNHPDAGGF 402

```

```

Qy 395 FKALYFANEMDSTRPVVMSMDADDETRDVALKPDYICVNRVYGYWYQRIEGL 454
    I : : : : :
Db 403 IRREVELFKSLDSSRPVTFAS-----HRSVRDLALAEYDVIVSLNYHGMWEMGDIDSGV 457
    : : : : :
Qy 455 QALEKOLEEYLAH-RKPTFVTFEGDALAGIHYPDPQMFSESEYQALVEVTITLLKKD 513
    : : : : :
Db 458 KVAIAIELEEIHKKFPEKPIIITEFGDALYGHSDPPQMWSESEYOSEMIKRYIALEAREKD 517
    : : : : :
Qy 514 YIICTHWAFADKTPQNVRRPILNHKGFTPDQPKLVAVYLRRLM 560
    I : : : : :
Db 518 YIYGFHLMWADRFTRPQNSRITLNNKGFFTDRDQPKLAAKAVLELF 564
    : : : : :

```

```

RESULT 6
US-10-120-145-2
/ Sequence 2, Application US/10120145
/ Publication No. US20030157684A1
/ GENERAL INFORMATION:
/ APPLICANT: Jefferson, Richard A.
/ APPLICANT: Kiljan, Andrzej
/ APPLICANT: Keese, Paul Konrad
/ TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
/ TITLE OF INVENTION: USES THEREOF
/ FILE REFERENCE: 190106.405
/ CURRENT APPLICATION NUMBER: US/10/120,145
/ CURRENT FILING DATE: 2002-04-11
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
/ NUMBER OF SEQ ID NOS: 71
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2
/ LENGTH: 602
/ TYPE: PRT
/ ORGANISM: Bacillus sp.
US-10-120-145-2

```

	Query Match	31.0%;	Score 929;	DB 14;	Length 602;	
	Best Local Similarity	35.6%;	Pred. No. 1.7e-68;			
	Matches 221;	Conservative 93;	Mismatches 224;	Indels 82;	Gaps 15;	
QY	1 MVRPQRKKRFFILLGWNLEV-----TSKRPLAVPSNME--QYDDLC 44	:	:	:	:	:
Db	1 MLVPINTETRGVDINGWNNFKLDYGKGLSEKWYSKLTDTISMAVPSYNDIGVTYEIR 60	:	:	:	:	:
QY	45 YEGRPFYKTFPVPPXKSOKHRLRYFAALVNTDCEVFNGKVGNHELYLPEFVDATGK 104	:	:	:	:	:
DQ	61 NHIGVYVEEFPVPAVLKDQRIVLRFSGATIKALVIYNGELVNHKGGFIPFEALINNS 120	:	:	:	:	:
QY	105 VKGENELRVVENRLKVGSPSKYPDSCGTHTVGEFGS-----FRPNDFEPFYG 154	:	:	:	:	:
DQ	121 LRQGMNRVTVAVDNIL-----DSDTLRVGLYSRHEGLGKVIANKNPDEFNYA 170	:	:	:	:	:
QY	155 GTRRPVLIEETDHARLLDIWDTSBSEPKKLGKVKVGIENVEEAIVGEMTIKLG--EEB 212	:	:	:	:	:
DQ	171 GLHRPVIKIYTPPTVEDISVTDFTNGCP---TGTVTYIVDPQ----GRAETVKVSVDEE 223	:	:	:	:	:
QY	213 KKIRTSNRPFEGEFILLENARFMSLBEPRLYLPLKVLBEDDTLDI-----GIRLSIDEX 267	:	:	:	:	:
DQ	224 GKVASAEESGSNVLEIPNLILWEIRLNTYYOIKVELVNDGLTIDYEPFQVRRLVEVNDG 283	:	:	:	:	:
QY	268 RLVLNGRPVLLKFGSGHEEPFLVGOSTFYRLMIKOFNLKMIINNPSFTSYRYSEEMLD 327	:	:	:	:	:
DQ	284 KFLINKKPFLPKFGKHEDPIINGKSGFNASVMDFNLTKMIGANSFRTAIAPYSEELMR 343	:	:	:	:	:
QY	328 LADRLGLIVDEAPHNGIIRRYHN-----PETOKIA-----EDNIIRMIDR 368	:	:	:	:	:
DQ	344 LADREGLVVIDERPAVG---HNHPATTGLGEBSERSITWEKIRTEPHNDQVLRELVS 400	:	:	:	:	:
QY	369 HKQHPVIMSVANPEFSNHPDAEGFFKALYETANEKD--RTRPVVVWSMDARDERTDV 427	:	:	:	:	:
DQ	401 DKHNPSSVMMISINEAATEEGAYEKVFLVLTGYELDRPORRPTIYLVFWATGE--TDK 458	:	:	:	:	:

```

QY      428  ALKTYDIYCVNRYYGMYIYQGIIEBGLQALBEKDIEELYAKHR---KPIFVTEFGADATA 483
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      459  VAEIIDIATLNAYNNMYEFGDGLAEAKYLR---QEFHAMNRKCPSPKIMITEYAGDTVA 515
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      484  GIHYDPQMFSEHYQALVEKTRILLKKDYIIIGTVAFADPKTPQNNVRRPILNHGVF 543
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      516  GFHDIDIPMFITEEYQVEYYQANHVVDEFENFVGEQANMFADPATISQVNRVQNNKGVF 575
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      544  TRDRQPKLAHYLRRLMSFV 563
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      576  TRDRQPKLAHYFRERWTNI 595
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-10-364-649-2
; Sequence 2, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge B. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCONIDASE GENES,
; FILE REFERENCE: 190106.405C1
; CURRENT FILING DATE: 2003-02-12
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-2

```

Query Match	31.0%	Score 929,	DB 15,	Length 602,
Best Local Similarity	35.6%	Pred. No. 1,7e-68,		
Matches 221,	Conservative 93,	Mismatches 224,	Indels 82,	Gaps 15
QY	1	MVRPORNKKRILLINGVWNLGV-----TSKDRPIAVPSGSMNE--QYDLC 44		:::
Db	1	MLVPIINTEIRGVPLDINGVWNLKDYQKGLBEKWTESKLTIDTISMAVPSGYNDIGTKEIR 60		:::
QY	45	YEEGFTYKTTFFYPFKLSQKHIRLYFAAVNTDCEVFLNGEKVGNHIEYLPEVDVTGK 104		:::
Db	61	NHIGVWVEREFTVPAYLKQRIYLRFSSATHKQILVYVNGELVYEHKGFLPFEKEINNS 120		:::
QY	105	VKSGENELRVVVENRLKVGCFPSKVPDSCGTVTGFCGS-----FPANPDPFPFG 154		:::
Db	121	LRDGNKRTAVAVDNL-----DSTLIPVGLISERHEGLQKVIIRNKNPFPFFNYA 170		:::
QY	155	GIIRPVLIIEFDHARLIDIVDTSSEBEPEKKGVKKVILEVSEAVGQEMTIKLG--EE 212		:::
Db	171	GLHHPVKIYTPPTFYVEDISVATVDFNGP---TGTVTVYVDQ---GAEITVAVSVVDE 223		:::
QY	213	KKIRTSNRFVEGSEFLENAFMSLEDPVLPYKVLKEDQETLID-----GRTISWDEK 267		:::
Db	224	GKVAVSTGSLGQNVIEIPVILMEPIANTLYIQIKELVNDGLTIDVYEEBPGVRIYVANDG 283		:::
QY	268	RLVYNGKRVPLKGFQKHEEPVLQGTFFPLMIKDFNLKMINANSFRTSHYPSYSEWLD 327		:::
Db	284	KFLINNKFPYFKGCKHEHDPIINGRGFEASNVDMFNILKIGANSFPTAHYPSYSEELMR 343		:::
QY	328	LADRLGLIVIDEAPHVIGITRYHN-----PEQOKA-----ENIRMRIDR 368		:::
Db	344	LADREGVLVIDETPAVGV---HLNFMATTCGSGSERVSTWEKITEFHQCVLDELVS 400		:::
QY	369	HKNHPSVIMSVANEPESNHPDAGGFFKALYETANEMD--RTRPVVMSGMDAPDERTDV 427		:::
Db	401	DKNHPVVMGSIANAATAEEGAAVEYFPLVELTEKIDPOKRPVITVYFVMAATP--TDK 458		:::

[illegible]

RESULT 8

```

US-10-364-649-8
; Sequence 8, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106, 405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-8

```

Query Match	31.0%	Score 929	DB 15	Length 602
Beet Local Similarity	35.6%	Pred. No. 1.7e-68		
Matches 221	Conservative 93	Mismatches 224	Indels 82	Gaps 15
Qy	1	MVRPORNKKRFLILNGVWNLV-----TSKDRPIVPGSMNE--QYDLC	44	
Db	1	MLYPIINETRGVFDLNGVWNEFKLDYDGKLEEKWYSEKLTDTISMPVSSYINDIGATKEIR	60	
Qy	45	YEEBFTTKTTFYVPKXLSOKHIRLRYPAAVNTDCVFLNGEKKGNHIEYLFEEVDVTGK	104	
Db	61	NHIGVWVERERTVPAVLKDKQRIVLRFSGATHKALVYVNGELVYEHKGFRLFPEALINNS	120	
Qy	105	VKSGENELVYVENRLKVCGPSPKVPDSGTHTVGFGS-----PPANFDFPYG	154	
Db	121	LRDGNRRTVAVDNLT-----DDSTLPVGLYSEHBEGLKVIIRNKNGFDFFNYA	170	
Qy	155	GIIRPVLIETFDHARILIDVNTDSESEBPKLGGKVKKILEVSEAVQDEMTIKG--EE	212	
Db	171	GLHRPVKLYTTFPYVEDISVTTDNGR---TGTYTVDYFQ---GKAEYKVASVDEE	223	
Qy	213	KKIRTSNFEVGEFILIENARFWSLEDPYLPYKVELKDEYVLDI-----GIRTISWDEK	267	
Db	224	GKVAVASTGSLGNNVEIPVILMEPIANTLYIQIKVELVNDGLTTIDYEEBPGRYVEVNDG	283	
Qy	268	RLYLNGKEVPLKGFQKHEEPVLAGOGTFYPLMIKQFNLLKWINASFRTHSPYSEWLD	327	
Db	284	KFLINNKCFYFKGFCGHEDPTINGSGFMASSVMDPNLLKWIIGANSFRTAHRYPEBELMR	343	
Qy	328	LADRLGILVIDAPRVGILTRHYN-----PETQXIA-----EDNIRMIDR	368	
Db	344	LADREGLVVIDETPAVGV---HLNMTATGLEGESERVSTWEKIRTFEHDQVLELVS	400	
Qy	369	HGNPSVIMSSVANEPSNHPDAGEFKALYETANEMD--RRPYYVSSMDAPDRERTDV	427	
Db	401	DKNRPSSVMSIANEATBEBCAGAYEFKPVLLTELDQKRPVITLVFWATPE--TDK	458	
Qy	428	ALKKFDIVCVNRYYGMVLYOGRIEGLQALKEIDIELYARNR---KPIFYVEGADIVA	483	

[illegible]

RESULT 9

```

US-10-364-649-15
; Sequence 15, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/0/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
;
; LENGTH: 602
;
; TYPE: PRT
;
; ORGANISM: Bacillus sp.
US-10-364-649-15

```

Query Match	31.0%	Score 929;	DB 15;	Length 602;
Base Local Similarity	35.6%	Fred. No. 1.7e-68;		
Matches 221;	Conservative 93;	Mismatches 224;	Indels 82;	Gaps 15
Qy	1	MVRPQNRKRFILLINGVWNLV-----TSKDRPIAVPSGWN--OYDLC	44	
Db	1	MLVYINETREGVPLDNGVWNLKDYGKLEEKWESKLTDTISNAVPSNDIGYTKER	60	
Qy	45	YEEBPFTYKTTFFVPPKLSQKHRLYEAAVNTDCEVFLNGEKVGNHLEVLPEVDVTK	104	
Db	61	NHIGVWYEREFYTPAYLKDQRIYLRFSGASTHKAIVYVNGELVWEHKGFPLFEALINNS	120	
Qy	105	VKSGENELRVVYENRLKVGFPSPKVPDPSGTHTVGFPGS-----FPANPFPFPYG	154	
Db	121	LRDMNRRTAAVANDIL-----DDSTLPVGLYSERHEEGAKVIRANKNPDPFNVA	170	
Qy	155	GIIRPVLIETFDHARILDIWDVTSSEBPEKKLGKVKVIEVSEAVGQEMTKLG--EE	212	
Db	171	GLHRPVKITPTTFPYVEDISVTDPNGP--TGIVTYIVDQ-----GAEIYVSVDE	223	
Qy	213	KKIRTSNRPEGEETLENARFWSLEDPLYLKYLEKDEYTLDI---GIRTSWDEK	267	
Db	224	GKVAVSTEGLSGANEIRPVILMEPLNLYXOIKELVNDGTLTIIVYEEBPFGRVTEVNDG	283	
Qy	268	RLVYNGRPVFLKFGKHEEPVULGOGTFYPLMIDPNLKKIINNPSFTSHYPSSEMLD	327	
Db	284	KELINNKPFYKGGKCKEDTPINRGFRVNASVNDFFNLKIKIGANSFTAIYPSSELMR	343	
Qy	328	LADRLGLIVIDEAPHVGITRHYN-----PETOKIA-----EDNIRRMIDR	368	
Db	344	LADREGLVIDETPAVGV--HLNFMAATTGEGSERVSTWEKIRTFEPHODVIRELVS	400	
Qy	369	HKHNPSTVMSVAPEPSNHPDAGCFKALYETANEMD--RRRPVYVMSMDAPBERTDV	427	
Db	401	DKNRPVWMSINAEAAITEEGAEYFPPVELTELKELDPQRRPVITIVLEVATPE--TDK	458	
Qy	428	ALKFEDIVCVRRYGMWYIYOGRIEGQALEKDEKIDTEELYARRH---KPIFTFEGADAIA	483	

Job time : 170 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 19:13:30 ; Search time 7488 Seconds
(without alignments) 10929.602 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atggtacagccgcaacgaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sbs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	6	AR428149 Sequence
2	1688	99.9	12583	1	AE001766 Thermotog
3	194	11.5	2153	4	AF012423 Felis cat
4	192.4	11.4	2014	4	AF012424 Felis cat
5	190.2	11.3	2199	4	AF019759 Canis fam
6	174.4	10.3	2155	9	AF084552 Chlorocob
7	174	10.3	2128	6	CO575703 Sequence
8	174	10.3	2182	9	BC014142 Homo sapi
9	174	10.3	2321	3	BT010073 Drosophi1
10	172.4	10.2	1956	6	AR344264 Sequence
11	172.4	10.2	1956	6	AX147652 Sequence
12	172.4	10.2	2191	6	CO847932 Sequence
13	172.4	10.2	2191	9	HUMGLCB
14	166.8	9.9	2391	10	RNGLCB
15	166.8	9.9	2472	6	AX827491 Rat mRNA fo
16	166.8	9.9	2472	10	RATGLCB
17	159.6	9.4	1947	6	AR344265 Sequence
18	159.6	9.4	1947	6	AX147654 Sequence
19	159.6	9.4	2456	10	MUSGUS

20	159.6	9.4	2456	10	MUSGUS	M28541 Mouse beta-
21	159.6	9.4	2456	10	MUSGUS	M63836 Mouse beta-
22	159.6	9.4	2492	10	BC071226	BC071226 Mus muscu
23	157.8	9.3	2454	10	MUSGLCGUS	M19279 Mouse beta-
24	156.4	9.3	2455	10	MUSGLC	J03047 Mouse beta-
25	154.8	9.2	2081	5	AJ720880	AJ720880 Gallus ga-
26	142.2	8.4	23078	6	CO363720	CO363720 Sequence
27	142.2	8.4	60265	1	AE017283_25	Continuation (26 o
28	141.4	8.4	1809	12	AF354047	AF354047 Synthetic
29	141.4	8.4	1887	6	AR210802	AR210802 Sequence
30	141.4	8.4	1887	6	BD074927	BD074927 Microbial
31	141.4	8.4	7018	12	AY452736	AY452736 Reporter
32	141.4	8.4	11846	12	AF354045	AF354045 Binary ve
33	141.4	8.4	11821	12	AF354046	AF354046 Binary ve
34	141.2	8.4	2235	11	BV175016	BV175016 sqm76123
35	139	8.2	10775	1	AE006894	AE006894 Sulfolobu
36	138.6	8.2	2352	3	BT012475	BT012475 Drosophi1
37	138.6	8.2	2445	3	AY069329	AY069329 Drosophi1
38	130.4	7.7	1888	6	AR428153	AR428153 Sequence
39	127.6	7.6	93590	3	AC084452	AC084452 Caenorhab
40	123.6	7.3	301450	1	AP003185	AP003185 Clostridi
41	122	7.2	3448	1	CPE420784	AJ420784 Clostridi
42	115.6	6.8	2105	1	AF305917	AF305917 Escherich
43	115.6	6.8	300330	1	AP005222	AP005222 Corynebac
44	114	6.7	1814	6	BD185061	BD185061 Nucleic a
45	114	6.7	2103	1	AF305918	AF305918 Escherich

ALIGNMENTS

RESULT 1	AR428149	Sequence 14 from patent US 6641996.	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428149	Sequence 14 from patent US 6641996.	DNA	linear	PAT 18-DEC-2003
DEFINITION	AR428149	Sequence 14 from patent US 6641996.	DNA	linear	PAT 18-DEC-2003
ACCESSION	AR428149	Sequence 14 from patent US 6641996.	DNA	linear	PAT 18-DEC-2003
VERSION	AR428149.1	GI:40187542	DNA	linear	PAT 18-DEC-2003
KEYWORDS	Unknown.		DNA	linear	PAT 18-DEC-2003
SOURCE	Unknown.		DNA	linear	PAT 18-DEC-2003
ORGANISM	Unknown.		DNA	linear	PAT 18-DEC-2003
REFERENCE	Unclassified.		DNA	linear	PAT 18-DEC-2003
AUTHORS	Jefferson, R.A. and Mayer, J.E.		DNA	linear	PAT 18-DEC-2003
TITLE	Microbial .beta.-glucuronidase genes, gene products and uses thereof		DNA	linear	PAT 18-DEC-2003
JOURNAL	Patent: US 6641996-A 14 04-NOV-2003;		DNA	linear	PAT 18-DEC-2003
FEATURES	Location/Qualifiers		DNA	linear	PAT 18-DEC-2003
source	1.1689		DNA	linear	PAT 18-DEC-2003
ORIGIN	/organism="unknown"		DNA	linear	PAT 18-DEC-2003
	/mol_type="genomic DNA"		DNA	linear	PAT 18-DEC-2003
Query Match	99.9%	Score 1688; DB 6; Length 1689;	DNA	linear	PAT 18-DEC-2003
Best Local Similarity	100.0%	Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;	DNA	linear	PAT 18-DEC-2003
Matches 1689; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		DNA	linear	PAT 18-DEC-2003
QY	1	ATGTTAGACCGCAACGAAAGAGATTATTTCTTATCTGATGAGATTGGAT	60		
DB	1	ATGTTAGACCGCAACGAAAGAGATTATTTCTTATCTGATGAGATTGGAT	60		
QY	61	CTTGAAGTACCGCAACGAAAGAGATTATTTCTTATCTGATGAGATTGGAT	120		
DB	61	CTTGAAGTACCGCAACGAAAGAGATTATTTCTTATCTGATGAGATTGGAT	120		
QY	121	CAGATCTGTGCTACGAGAGAGACCTTCACTCAAAACCACTTCACTTCCGAAG	180		
DB	121	CAGATCTGTGCTACGAGAGAGACCTTCACTCAAAACCACTTCACTTCCGAAG	180		
QY	181	NAACTTCAAAAACACATCAGACTTCTTCTTCTGCTGAGACGAGACTCGAGGTC	240		
DB	181	NAACTTCAAAAACACATCAGACTTCTTCTTCTGCTGAGACGAGACTCGAGGTC	240		
QY	241	TTCTCAACGAGAGAAAGTGGAGAGATCATGATTAATCTTCTTCTGAGTAAT	300		

```

Db      |||
241 TTCTCAACGAGAGAAAGTGGAGAGATCAATTGAATACCTTCCCTCGAAGTAGAT 300
Qy      |||
301 GTGACGGGGAAAGTGAATCCCGAGAGAAACGAATCAGGAGTGTGTTGAGAAACGATTG 360
Db      |||
301 GTGACGGGGAAAGTGAATCCCGAGAGAAACGAATCAGGAGTGTGTTGAGAAACGATTG 360
Qy      |||
361 AAAGTGGAGAGATTTCCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGGATTTT 420
Db      |||
361 AAAGTGGAGAGATTTCCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGGATTTT 420
Qy      |||
421 GGAAGTTTTTCCACTGCAAACTTTCGACTTTCCCTACGAGTGAATCAAGGCTGT 480
Db      |||
421 GGAAGTTTTTCCACTGCAAACTTTCGACTTTCCCTACGAGTGAATCAAGGCTGT 480
Qy      |||
481 CTGATAGATTTGACAGACCGAGAGATCTCGACATTTGGTGGAGACAGAGTAGTCT 540
Db      |||
481 CTGATAGATTTGACAGACCGAGAGATCTCGACATTTGGTGGAGACAGAGTAGTCT 540
Qy      |||
541 GAAACGGAGAGAACTTGGAAAAGTGAAGATGAGAGTCTCAGAAAGACGGTG 600
Db      |||
541 GAAACGGAGAGAACTTGGAAAAGTGAAGATGAGAGTCTCAGAAAGACGGTG 600
Qy      |||
601 GGAAGAGAGATGACATCAAACTTGGAGAGAGAGAGAGAGATTAAGAACATCCACAG 660
Db      |||
601 GGAAGAGAGATGACATCAAACTTGGAGAGAGAGAGAGAGATTAAGAACATCCACAG 660
Qy      |||
661 TTCCGTGAGAGGGAGATTCCTCGAAAAACCGCAAGTTTGGAGGCTCGAAGATCCATAT 720
Db      |||
661 TTCCGTGAGAGGGAGATTCCTCGAAAAACCGCAAGTTTGGAGGCTCGAAGATCCATAT 720
Qy      |||
721 CTTTATCTCTCAAGGTGGAACCTTGAAGAAACGAGTACATCTGAGACCTCGAATCAGA 780
Db      |||
721 CTTTATCTCTCAAGGTGGAACCTTGAAGAAACGAGTACATCTGAGACCTCGAATCAGA 780
Qy      |||
781 ACGATACGCTGGAGACGAGAAAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGC 840
Db      |||
781 ACGATACGCTGGAGACGAGAAAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGC 840
Qy      |||
841 TTTGGAAGACAGAGAAATCCCGTTCGGGGACGGGACCTTTTATCCATGATGATA 900
Db      |||
841 TTTGGAAGACAGAGAAATCCCGTTCGGGGACGGGACCTTTTATCCATGATGATA 900
Qy      |||
901 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAAATCTTTGAGACCTCTCATATCT 960
Db      |||
901 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAAATCTTTGAGACCTCTCATATCT 960
Qy      |||
961 TACAGTGAAGAGTGGCTGGATCTTTCGACAGACTTCGGAATCTTGTGATAGACGAAGCC 1020
Db      |||
961 TACAGTGAAGAGTGGCTGGATCTTTCGACAGACTTCGGAATCTTGTGATAGACGAAGCC 1020
Qy      |||
1021 CGGACGTTGGTATCAACAGGTAACCACTTCCCGAGACTCAGAAAGTATACAGAAAGC 1080
Db      |||
1021 CGGACGTTGGTATCAACAGGTAACCACTTCCCGAGACTCAGAAAGTATACAGAAAGC 1080
Qy      |||
1081 AACATTAAGAAAGATGATCGACAGACCAAGAACCAATCCAGTGTGATCTGAGAGTGTG 1140
Db      |||
1081 AACATTAAGAAAGATGATCGACAGACCAAGAACCAATCCAGTGTGATCTGAGAGTGTG 1140
Qy      |||
1141 GCGAACGAACCAAGAGTCCAAACATCCAGACGCGGAGGGTGTCTTCAAAAGCCCTTATGAG 1200
Db      |||
1141 GCGAACGAACCAAGAGTCCAAACATCCAGACGCGGAGGGTGTCTTCAAAAGCCCTTATGAG 1200
Qy      |||
1201 ACTGCAATGAAGATGATGAAACAGCGCCCGTTCATAGTGTGATGATGATGAGACCAACA 1260
Db      |||
1201 ACTGCAATGAAGATGATGAAACAGCGCCCGTTCATAGTGTGATGATGATGAGACCAACA 1260
Qy      |||
1261 GACGAGAGAACAGAGACGTCGCTGAAGTACTTTCGACATGCTGTGTGAACAGTATC 1320
Db      |||
1261 GACGAGAGAACAGAGACGTCGCTGAAGTACTTTCGACATGCTGTGTGAACAGTATC 1320
Qy      |||
1321 TACGCTGTATCATCTATCAGGAGAGAGATGAGAAAGGATTCAGAGCTTGGAAAAAGAC 1380

```

```

Db      |||
1321 TACGCTGTATCATCTATCAGGAGAGATGAGAAAGGATCTTCAAGCTCTGAAAAAGAC 1380
Qy      |||
1381 ATGAGAGGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAGAGATTTGGTCGGAC 1440
Db      |||
1381 ATGAGAGGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAGAGATTTGGTCGGAC 1440
Qy      |||
1441 GCGATAGCTGGCATCCACTACATCAGATCCACTTCAAAATGTTCTCCGAGAGATCAAGCAG 1500
Db      |||
1441 GCGATAGCTGGCATCCACTACATCAGATCCACTTCAAAATGTTCTCCGAGAGATCAAGCAG 1500
Qy      |||
1501 CTGCTTGAAGAAAGAGATCAGGCTCTTTGAAAAAGACTACATCATCTGAAACACAGTG 1560
Db      |||
1501 CTGCTTGAAGAAAGAGATCAGGCTCTTTGAAAAAGACTACATCATCTGAAACACAGTG 1560
Qy      |||
1561 TGGGCTTTGAGATTTTAAAGCTCTCAGATGAGAGAGAGAGAGAGATTCACACAGCAG 1620
Db      |||
1561 TGGGCTTTGAGATTTTAAAGCTCTCAGATGAGAGAGAGAGAGAGATTCACACAGCAG 1620
Qy      |||
1621 GGTGTTTTCAAGAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
Db      |||
1621 GGTGTTTTCAAGAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
Qy      |||
1681 AGTGAAGTT 1689
Db      |||
1681 AGTGAAGTT 1689

RESULT 2
AB001766/c 12583 bp DNA linear BCT 04-JUN-2004
LOCUS
DEFINITION Thermotoga maritima MSB8 section 78 of 136 of the complete genome.
ACCESSION AB001766 AB000512
VERSION
KEYWORDS
SOURCE
ORGANISM Thermotoga maritima MSB8
Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
REFERENCE
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., Eisen,J.A., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316
PUBMED 10360571
REFERENCE 2 (bases 1 to 12583)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
FEATURES
source
1..12583
location/Qualifiers
1..12583
organism="Thermotoga maritima MSB8"
mol_type="genomic DNA"
strain="MSB8"
db_xref="taxon:243274"
81..1607
locus_tag="TM1058"
81..1607
locus_tag="TM1058"
/note="Similar to GB:AB000666 percent identity: 48.70; identified by sequence similarity; putative"
/codon_start=1

```

```
gene
/translate=1
/product="glutamate synthase-related protein"
/protein_id="AAD36135.1"
/db_xref="GI:4981601"
/translation="MSGLCACVDCGTCGCTWLASFRGEVLYPGPFGDITAGAVKD
YVDVSHNITLICYARGAEGLPBEGPPTALTFTVVDITTEYGMIDIKVMKPIPTGA
LGSITARKNMHIAVGAALISGITVVCENAVGVDPLEDLSNGKVSPELDRIEI
YKRVHDEGGEITLQMVNEDRLGVAEYVINGGIEETELKMGCAKSIIGETIKRSL
BRALFKRGYIVLPDPELPYQAPKSEGIKEFRHRSLGCVSKESFTKEVERIKL
GKRITLTGAYSAAVLAALRYGAARVDLTVDGAPGCTGMSMPMNENGIPTFY
LEALTYQPAEKLSRGIRVDPDIAAGSETEGVKAIAMGSPYKACMGKALMIPA
MWKNIGELKSGNPKTVSKYGTVEEFTVEELRSRFGEEVKKLPJLGIAYTE
VQFKTGLQQLMAARKFRSLASRKDILATKDAEISGIPYVWESYRDEARLLEE
"
/complement(1650..1913)
/locus_tag="TM1059"
/complement(1650..1913)
/locus_tag="TM1059"
/notes="similar to SP:P45693 PID:862985 GB:AL009126 percent
identity: 48.70; identified by sequence similarity;
putative"
/codon_start=1
/translate=1
/product="spovs-related protein"
/protein_id="AAD36136.1"
/db_xref="GI:4981602"
/translation="WEVLKVSCKSDPNKVAIAAGVREHGAETIOAIGAGVNOAVK
AIAARGYLAPEGIDLVFPAFTDVEIENEKRTAKFVFPKS"
/complement(1962..3272)
/locus_tag="TM1060"
/complement(1962..3272)
/locus_tag="TM1060"
/notes="similar to GB:AE00792 percent identity: 49.74;
identified by sequence similarity; putative"
/codon_start=1
/translate=1
/product="conserved hypothetical protein"
/protein_id="AAD36144.1"
/db_xref="GI:4981610"
/translation="WNISIEGALAVLINQFPGAYLTGYFLMGASSFFIGLPGSIP
FLANTLQTLTTSFHSRLKSRQIIVPLMTARTSILLFAVPAIKGILLAVLYFY
IOAGLSAPLQMSMSDLVQDMISYGFGRLLIGAVQIPAMPLAGAILDLSGNMK
GGGTLFIAGSLGALNGVYETKQVEPYKPRPASVITKAVYKLLKEHFKNFTLGP
FMNPAIGCTGYINMLKEVEFSYLOISLVNAVGMFIGTLFQPFMGKIDRYGQYF
LKVCIMHAIIVILMTLTPRSFLVFFLOIIGITVAGTSQLVFTYLMYAPSSLKT
SAFVSNSISNLSFAGSLVASVLVASLSENISLPGISAIRLTMFISFPLRASAAVYII
SRMDLGTPOKVDLSIOAVKESFFSGTWPWIRERLNTLNIFFRKR"
/complement(3382..4545)
/locus_tag="TM1061"
/complement(3382..4545)
/locus_tag="TM1061"
/notes="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/translate=1
/product="hypothetical protein"
/protein_id="AAD36137.1"
/db_xref="GI:4981603"
/translation="MRSMTDVRVYWSILSRKICEWPLELCASDRLLKSKMPYEGKKEE
RRKFTHELIGRIFCGISPLELNRESPTDPERKIATRLSEPAVKSLDVAUNNCK
DYNMKEGROPYDAAPLVEALIRAPKVLMBELDSTKRLRLKATRKIKPEYISNM
LIFSAMISITFFPAGESDSTKVDILKNEBSVYGDAVGGPFRRNDYKSFYTP
MTVDLARIISEKTSWKELYVVLRAARAVFARVENSPESTFPIIGRSITYRYAVF
HLISQLSLHLPLASISPAQVACALTAVLRIRFENPSTFDEMGKLIKIVIGSPGLGE
EYITGSLYLCTTVFLPLGLPTSDPFMDPCKMTNKKVGEDEVADPRALED"
/complement(4542..6233)
/locus_tag="TM1062"
/complement(4542..6233)
/locus_tag="TM1062"
/notes="similar to PID:642973 PID:14446 PID:412358 percent
identity: 57.73; identified by sequence similarity;
putative"
/codon_start=1
```

```
gene
/translate=1
/product="beta-glucuronidase"
/protein_id="AAD36143.1"
/db_xref="GI:4981609"
/translation="WVRPQRKKRPIILINGVNLVTSKDRPIAVPSMEOYODLC
YEEGPTATITFYVYKELSOQHILRYFAAVNTDCVFNLNGEYGNHILEYLFEDVT
GKVKSGEHLVAVENRLKVGSPKVDSGNHTVGFSPGPAPAPDFPPYGGILRPV
LIEPTDARILDIWVDSRPEKKLGVKXIKVYSERAVQGMETIKIGEEKXIRTS
NRFNGEITLENARWSLEDPYLYPLKYLEKDEYTLIDITRTISWDEKRLYLNKPV
FLVGGKREEFPVLQGGTFYPLMKDFMLKMLINANSRTSHYPYSEEWLIDLADRLGI
LVIIDAPVIGITRYHYNPETQKIEDNIRMI DRHKHPSYIMSVANPESHNDAB
GPFKALYETANEMDSTRPVVWMSMDADDERDVALKYFDIVCNARYGWTYOGRI
BEGQALKEQDIEIYARHKDIPVTEPADYIAGIHYPPOPMSESEYQAEIYKTRIL
LLKXDYITGTYVMAFADPKTPQNRPRPILNKGVTTRDQRLVAHVLRRLSEV"
/complement(6284..7309)
/locus_tag="TM1063"
/complement(6284..7309)
/locus_tag="TM1063"
/notes="similar to SP:P42065 PID:677944 GB:AL009126 percent
identity: 75.31; identified by sequence similarity;
putative"
/codon_start=1
/translate=1
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36138.1"
/db_xref="GI:4981604"
/translation="MALLEVNLKKYRPPIIKRFRKRVGYLKAVDGISPHIEGETL
GLVSGCGGKTTAKLIMRGIEPTEGEVILNMDGKVDITKLSEKELREKGVRLQIM
IFODYSLNRPVTRDIASPLVNGVYKKEEIDAVSLDLAVGALREKVMORPOM
APSGGQRORIAAVAILAKPKLVCDPETSALDVSVOQIINILKDKLQKENVLTFLFI
SHDQGVENITNRVAVMYGRIVEALTEBEPSSPKHYTTEALLSAVCPDPKPKRK
AQLTGEVPDPNSPBGCFHPRCPYAKAIKCEYVPLNVTGDPNHLVACHFADSLK
LEGVGTM"
/complement(7236..8294)
/locus_tag="TM1064"
/complement(7236..8294)
/locus_tag="TM1064"
/notes="similar to SP:P42064 PID:677943 GB:AL009126 percent
identity: 72.95; identified by sequence similarity;
putative"
/codon_start=1
/translate=1
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36139.1"
/db_xref="GI:4981605"
/translation="WSTLLOIKNRTYFTDGVYKAVDGVSEIEEGRTIGVSGSG
CGKSYTARSITIKLSTAGRIVSGELINMDQMDLVKFSKEIRKRGKRIAMIFQE
PMAFSPVYITGDQITBGMITYHFGITKQEARAVELLRVGIKPRPKMIDSYPEYS
GMRQORAMIAALSCNPRLLIADPPTNLDVYTIQAVYLDLKDLQOEYKMAIMMTN
MGVAEMADHVAVVWYGLVVSAPVEBELFVNPKHYSILHRSIPVVGKRYERLVEIE
GDVDPKRMPCGCFHPRCPYMMKGIQDERPEVEVEGPEHRVSCFLYGGEDGAS"
/complement(8306..9409)
/locus_tag="TM1065"
/complement(8306..9409)
/locus_tag="TM1065"
/notes="similar to GB:AE006657 percent identity: 67.80;
identified by sequence similarity; putative"
/codon_start=1
/translate=1
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36140.1"
/db_xref="GI:4981606"
/translation="MKKREOMKEKRYVASOMQLIMWRPRKRLAVIGSVLLIIVFA
ICRFPAPVDNKNKYRVYPAAPORIHFEHSGKGPVGVYATVTDLETARRIKED
KSKIFKIFPVRGDEYKRWGKTNINHTIGVDEDMFLIGDSIGRMDLSITTYGARI
STSIGLGVFLFIIIGVAIIGGISGFGGAVNFPQRTIEILIKSIPTPLWALSAPL
QNMPLRVRVYFVILSLTGWTDLARVVRSLRLREDFPMAKFMKASBARLIIFRH
MLP2FMHSLIASITLSIPGMIDIGETSFLIGLRPVIISWGVLLQEAQNTTVVALYP
MLIPVVFVITTVCFNPFVGGDLADADPVANM"
/complement(9420..10403)
/locus_tag="TM1066"
```

CDS

complement(9420..10403)
 /locus_tag="TM1066"
 /note="similar to SP:P42062 PID:677946 GB:AL009126 percent identity: 62.93; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="oligopeptide ABC transporter, permease protein"
 /protein_id="AAD36141.1"
 /db_xref="GI:4981607"

Query Match 99.9%; Score 1688; DB 1; Length 12583;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ATGGTAAGACCGCAACGAAAGAAAGAGATTATTTCTTATCTTGAATGGAATTTGGAT 60
DB 6233 ATGGTAAGACCGCAACGAAAGAAAGAGATTATTTCTTATCTTGAATGGAATTTGGAT 6174
QY 61 CTGTAAGTAACGACGAAAGACAGACCAATGCGCGTTCTGGAAGCTGGAATGAGCATAC 120
DB 6173 CTGTAAGTAACGACGAAAGACAGACCAATGCGCGTTCTGGAAGCTGGAATGAGCATAC 6114
QY 121 CAGGATCTGTGCTACGAGAAGAGACCTTACCTTACCAAAACCACTTCTTACGTTCCGAG 180
DB 6113 CAGGATCTGTGCTACGAGAAGAGACCTTACCTTACCAAAACCACTTCTTACGTTCCGAG 6054
QY 181 NAACTTTTCACAAAACATCAGACTTTTACTTTGCTGCGGTGGAACAGGACCTGCGAGGTC 240
DB 6053 NAACTTTTCACAAAACATCAGACTTTTACTTTGCTGCGGTGGAACAGGACCTGCGAGGTC 5994
QY 241 TTCTCTCAACGAGAGAAAGTGGAGAGATCATATTGAATATACCTTCCCTTCGAAATGAT 300
DB 5993 TTCTCTCAACGAGAGAAAGTGGAGAGATCATATTGAATATACCTTCCCTTCGAAATGAT 5934
QY 301 GTGACGGGGAAAAGTGAATCCGAGAGAAACGAACTCAGGGGTGTTGTTGAGAACAGATTG 360
DB 5933 GTGACGGGGAAAAGTGAATCCGAGAGAAACGAACTCAGGGGTGTTGTTGAGAACAGATTG 5874
QY 361 AAAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGACCTCACACCGTGGGATTTTTT 420
DB 5873 AAAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGACCTCACACCGTGGGATTTTTT 5814
QY 421 GGAAGTTTTTCACCTGCAAACTTTCGACTTCTTCCCTACGATGATCATTAAGGCTGTT 480
DB 5813 GGAAGTTTTTCACCTGCAAACTTTCGACTTCTTCCCTACGATGATCATTAAGGCTGTT 5754
QY 481 CTGATAGAGTTCAACAGACCGAGAGATATCTGACATCTGGGTGACACGAGTAGTCT 540
DB 5753 CTGATAGAGTTCAACAGACCGAGAGATATCTGACATCTGGGTGACACGAGTAGTCT 5694
QY 541 GAAACGGGAGAAAGAACTTGGAAAGTGAAGTGAAGTGAAGTCTCGAAGAAAGCGGTG 600
DB 5693 GAAACGGGAGAAAGAACTTGGAAAGTGAAGTGAAGTGAAGTCTCGAAGAAAGCGGTG 5634
QY 601 GACACGAGATGACGATCAAACTTGGAGAGAGAGAAAGAAAGATTTAGAACATCCAACAGA 660
DB 5633 GACACGAGATGACGATCAAACTTGGAGAGAGAGAGAAAGAAAGATTTAGAACATCCAACAGA 5574
QY 661 TTTCGTGAAGGGGAGTTTCTCTCGAAAACGCCAGGTTCTGAGGCTTCGAAGATCCATAT 720
DB 5573 TTTCGTGAAGGGGAGTTTCTCTCGAAAACGCCAGGTTCTGAGGCTTCGAAGATCCATAT 5514
QY 721 CTTTATCCCTCAAGGTGGAACCTTGAAGAGACGAGTACATCTGAGACATCGGAATCGA 780
DB 5513 CTTTATCCCTCAAGGTGGAACCTTGAAGAGACGAGTACATCTGAGACATCGGAATCGA 5454
QY 781 ACGATCAGCTGGGACGAGAGAGAGCTCTATCTGAACCGGAAACCTGCTTTTGAAGGAC 840
DB 5453 ACGATCAGCTGGGACGAGAGAGAGCTCTATCTGAACCGGAAACCTGCTTTTGAAGGAC 5394
QY 841 TTTGAAAACGACGAGGAATTTCCCGCTTCTGGGACAGGGGCACTTTTATTCATTTGATATA 900
DB 841 TTTGAAAACGACGAGGAATTTCCCGCTTCTGGGACAGGGGCACTTTTATTCATTTGATATA 900

```

```

DB 5393 TTTGAAAACGACGAGGAATTTCCCGCTTCTGGGACAGGGGCACTTTTATTCATTTGATATA 5334
QY 901 AAAAGCTTCAACCTTCTGGAAGTGAATCAACGCGAATTTCTTCAGAGCTTCACATACCT 960
DB 5333 AAAAGCTTCAACCTTCTGGAAGTGAATCAACGCGAATTTCTTCAGAGCTTCACATACCT 5274
QY 961 TACAGTGAAGAGTGGCTGGATCTTTCGCGACAGACTCGGAATCTTTGTATGACGAAACC 1020
DB 5273 TACAGTGAAGAGTGGCTGGATCTTTCGCGACAGACTCGGAATCTTTGTATGACGAAACC 5214
QY 1021 CCGACACTTGGTATTCACAGGTACCACTTCAATCCCGAGACTCAGAAAGATGACGAAAGAC 1080
DB 5213 CCGACACTTGGTATTCACAGGTACCACTTCAATCCCGAGACTCAGAAAGATGACGAAAGAC 5154
QY 1081 AACATTAAGAAATGATGACAGACACAAAGAACATCCAGTGTGATGATGAGAGGTG 1140
DB 5153 AACATTAAGAAATGATGACAGACACAAAGAACATCCAGTGTGATGATGAGAGGTG 5094
QY 1141 GCGAACGAAACCGAGTCCAACTCAGACCGCGAGGGTTTCTTCAAGCCCTTTATGAG 1200
DB 5093 GCGAACGAAACCGAGTCCAACTCAGACCGCGAGGGTTTCTTCAAGCCCTTTATGAG 5034
QY 1201 ACTGCAATGAATGATCCAGACACGCCCGCTTGTCTATGTGAGATGATGAGCGACCA 1260
DB 5033 ACTGCAATGAATGATCCAGACACGCCCGCTTGTCTATGTGAGATGATGAGCGACCA 4974
QY 1261 GACGAGAGAACAGAGAGCTGGCGCTGAAGTCTTCCAGTGTGTTGAACAGGTAC 1320
DB 4973 GACGAGAGAACAGAGAGCTGGCGCTGAAGTCTTCCAGTGTGTTGAACAGGTAC 4914
QY 1321 TACGCTGTGTAATCTATCAGGAGAAAGTGAAGAGACCTTCAAGCTCTGAAAAGAAC 1380
DB 4913 TACGCTGTGTAATCTATCAGGAGAAAGTGAAGAGACCTTCAAGCTCTGAAAAGAAC 4854
QY 1381 ATAGAAGACTCTTATGCAAGGACACAGAAAGCCATCTTTGTCAAGAAATTTGGTGGAC 1440
DB 4853 ATAGAAGACTCTTATGCAAGGACACAGAAAGCCATCTTTGTCAAGAAATTTGGTGGAC 4794
QY 1441 GCGATAGCTGGACATCCACTAGATCCACTCAAAATGTTCTCGAAGGTACCAAGCAGAG 1500
DB 4793 GCGATAGCTGGACATCCACTAGATCCACTCAAAATGTTCTCGAAGGTACCAAGCAGAG 4734
QY 1501 CTCGTTGAAAAGAGATCAGGCTCTCTTTGAAAAAGACTACATATCGAACAACAGT 1560
DB 4733 CTCGTTGAAAAGAGATCAGGCTCTCTTTGAAAAAGACTACATATCGAACAACAGT 4674
QY 1561 TGGGCTTTTGAAGATTTTAAAGCTCTCAGAAATGTGAAGAACCCATTTCTCAACCAAG 1620
DB 4673 TGGGCTTTTGAAGATTTTAAAGCTCTCAGAAATGTGAAGAACCCATTTCTCAACCAAG 4614
QY 1621 GGTGTTTTCAACAAGACAGACCAACCACTCGTGTCTATGTAGAGAAAGCTGTGG 1680
DB 4613 GGTGTTTTCAACAAGACAGACCAACCACTCGTGTCTATGTAGAGAAAGCTGTGG 4554
QY 1681 AGTGAAGTT 1689
DB 4553 AGTGAAGTT 4545

```

RESULT 3
 AF012423
 LOCUS AF012423 2153 bp mRNA linear MAM 09-SEP-1999
 DEFINITION Felis catus beta-glucuronidase (gusb) mRNA, complete cds.
 ACCESSION AF012423
 VERSION AF012423.1 GI:4102550
 KEYWORDS
 SOURCE
 ORGANISM
 Felis catus (cat)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Placentalia; Felidae; Felis.
 REFERENCE
 1 (bases 1 to 2153)
 Pyle,J.C., Kurzhals,R.L., Laessle,M.E., Henthorn,P.S., Alur,P.R.,
 Wang,P., Wolfe,J.H., Giger,U., Haskins,M.E., Patterson,D.F.,

QY	751	GACAGTACACTCTGTGATCGATCGAATACGAAGATCAGCTGGGAAGGAGGCTTAT	810
Db	972	GACTTCTACACTCTCTCCCGTGGGGATTGGACAGGTGGCTGTACAGAGCACCAATTTCTTC	1033
QY	811	CTGAACGGGAAAACCTGTCTTTTGGAGGGCTTTGGAAAGCACGAGGAATTCGCCGTTCG	870
Db	1032	ATCATATGGGAAACCTTTCTATTTCACG69GGTCAACAAGCACGAGGATTCAGATATCCGA	1092
QY	871	GGGACGGGACACTTTTATCCATGTATGTATTTAAAGACTTCAACCTTGAATGTATCAAC	930
Db	1092	GGGAAAGGCTTTGACTGGCAGCTGTGTGGAAAGACTTCAATTTCTTCGCTGGCTCGGG	1153
QY	931	GCGAATTTTTCAGGACCTCTCATTACTTCTTACATGGAAGATGGCTGGATTTGCGGAC	990
Db	1152	GCCAAAGCCTTCCGGACAGATCATAACCCCTAAGGGAGAGGTATGAGCTCTGTGAC	1211
QY	991	AGACTCCGAATCTCTTGTGTATAGACGAAGCCCGCACGTTGTGTATCAAGAATTACCA---C	1047
Db	1212	CGTATATGGAGTCGTGTATCGACGACGAGTCCCGGTGTGGGCGATGTGTGTCCAGAGAC	1273
QY	1048	TACATCCCGAGACTCAGAAGATAGACAAAGACAATTAAGAAATGATCGACAGACAC	1107
Db	1272	TACAGCAACGTGTCTCTGCAGACCACTGAGAGTGTATGAGAGACTGTGTGGCAGGGAC	1333
QY	1108	AAGAACCTCCCAAGTGTATCATGTGAGTGTGGGAAGAACCAAGAGTCCAACCATCCA	1167

Db	1332	AAAGATCACCCGGCGTGTGGATGTGGTCTGTGGCCAAATGAGCCCGCTTCTCTGAAA	1391
Qy	1168	GACCGGAGGGGTTTCTTCCAAAGCCCTTTATGAGCTGCGCAATGAATGATCGAAACGCG	1227
Db	1392	CCTCCCGGTTACTACTCTTCAAGACCGCTAATGGCTCACAAAGCCTTGGATCCCTCCGG	1451
Qy	1228	CCGGTTTCATGTGAGACATGATGAGGACCCAGACAGAAACAAAGACGTGGCGCTG	1287
Db	1452	CCCGTGACCTTTGTGACCAACTCCAACTATGAAAG-----CAGACTGGGGCG	1499
Qy	1288	AAGTACTTCGACATCGTCTGTGTAACAGGTACTACGGCTGTACTCTATCAGGGAAG	1347
Db	1500	CCGTATGTGACGTCATCTGTGTGAATAGTTACTACTCTTGGTATCATGACTATGGTCAAC	1559
Qy	1348	ATAGAAGAAGGACTTCAGGCTCTGGAAAAAGACATAGAAGAGCTTATGCAAGGCACAGA	1407
Db	1560	ATGAGAGTGATTCAGCTGACGCTGGCAACCAAGTTTGAAACTGTATDAGAACTTATCAG	1619
Qy	1408	AAGCCCATCTTTGTACAGAAATTCGGTGGGACGGCATAGCTGGCACTACATACATCCA	1467
Db	1620	AAACCAATTAATCCAGACCGAGTACGGGACAGACACCAATTCAGAGGCTTTCACACAGAACCA	1679
Qy	1468	CCTCAATGTCTTCCGAAGATACCAAGCAGAGCTCGTGTAAAAAGACATCAGGCTCCTT	1527
Db	1680	CCTTGATGTGTACGCGGGAATATACCAAAAGGTGTGCTGAGCAGTATCAATTTGGTCTG	1739
Qy	1528	TTGAAA-----AAAGACTACATCATCGGAACAACACTGTGGGCTTTGAGATTTTAAAG	1581
Db	1740	GATGAAAAAGCAAAAGATATCGTGTGCGAGAGCTCATCTGGAACTTTGCCGATTTTATG	1799
Qy	1582	ACTTCTCAGAAATGTGAGAGAACCCATTTCTAACCAACAAGGATGTTTTTACAAAGACAGAGA	1641
Db	1800	ACTAACCACTCACCGCAGAGAGTATATGGGGAATTAATAAAGGGGATCTTCACGCGCAGAGA	1859
Qy	1642	CAACCCCAACTCGTTGCTCATGTACTGAGA	1671
Db	1860	CAACCAAAAGGGGACAGGCTTCTTTTGGCA	1889

FEATURES	source
LOCUS	AF012424
DEFINITION	AF012424 2014 bp mRNA linear MAM 09-SEP-1999
ACCESSION	Felis catus cell-type fibroblast mutant beta-glucuronidase (GUSB)
VERSION	mRNA, GUSB-MPS VII mutant allele, complete cds.
KEYWORDS	AF012424
SOURCE	AF012424.1 GI:4102552
ORGANISM	Felis catus (cat)
	Felis catus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE	1 (bases 1 to 2014)
AUTHORS	Eyfe, J.C., Kurzhals, R.L., Laessle, M.E., Henthorn, P.S., Alur, P.R., Wang, P., Wolfe, J.H., Giger, U., Haskins, M.E., Patterson, D.F., Sun, H., Jain, S. and Yuhki, N.
TITLE	Molecular basis of feline beta-glucuronidase deficiency: an animal model of mucopolysaccharidosis VII
JOURNAL	Genomics 58 (2), 121-128 (1999)
MEDLINE	99296826
PUBMED	10366443
REFERENCE	2 (bases 1 to 2014)
AUTHORS	Eyfe, J.C., Kurzhals, R.L. and Henthorn, P.S.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-1997) Microbiology, Michigan State University, 413 Gilmer Hall, East Lansing, MI 48824, USA
	Location/Qualifiers
	1..2014

```

    /organism="Felis catus"
    /mol_type="mRNA"
    /db_xref="taxon:9685"
    /cell_type="fibroblast"
    1..2014
    /gene="GUSB"

```


Query Match	11.3%;	Score 190.2;	DB 4;	Length 2199;
Best Local Similarity	52.8%;	Pred. No. 7.9e-40;		
Matches 492;	Conservative 0;	Mismatches 418;	Indels 21;	Gaps 3;
QY	750	AGACGAGTACACTCTGAGATCGGAATCAGAACATCACTGAGGACGAGAAAGAGCTCTTA	809	
DB	1010	AGACTTCTATCTACTCCCGTGGGATTTGGACCGTGGCGGTACAGAGGCGCAAGTTCTT	1069	
QY	810	TCGAAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGACGAGGAAATCCCGCTTCT	869	
DB	1070	CATCAACGGGAAACCTTTCTATTTTCCATGGGCTCAACAAACATGAGAGATGCCATATCCG	1129	
QY	870	GGGGCAGGACCTTTTATCCATGATGATTAAGACTTCAACCTTGAAGGATCA	929	
DB	1130	AGGAAAGGCTTTGACTGGCCGCTGCTGGTGAAGACTTCAACCTTGGCTGGG	1189	
QY	930	CGGAAATTTTTCAGAGCTCTCACTATCTTACAGTGAAGATGAGCTTGGCCGA	989	
DB	1190	CGCAATGCTTCCGACACGACCTACCCCTACGGGAGGAGTATGACAGCTTGGCA	1249	
QY	990	CAGACTCGAATCTTGTGATAGAGAGAGCCCGCACGTTGATACAAAGTACA	1046	
DB	1250	CGGCTATGGAGTCTGTGATCAGACGAGAGCCCTGGTGGCATCATGCTGGTCCAGAG	1309	
QY	1047	CTACAAATCCGAGACTCAGAGATAGCAGAGAACATTAAGAGATGATGACAGACA	1106	
DB	1310	CTACAGCAATGTGTCTCTGACAGACATCTGGAGTGAAGGAGAGCTGTGCTGGGA	1369	
QY	1107	CAGAACATCCAGTGTGATCATGTGAGTGGCGAAGAACAGAGTCCAAACCATCC	1166	
DB	1370	TAAAGATCAACCATCTGTAGTACATGTGTGTGATGAGCCACCTTCTCTGAA	1429	
QY	1167	AAGACGGAGGTTTTCAGAACCTTTATGAGACTGCCAATGAATGATGAAACAGC	1226	
DB	1430	GCCTGTGCTTACTACTTCAAGACGCTGATGCTCACACCAAGGCTTGGACCCCTCCG	1489	
QY	1227	CCCCGTGTCATGTGAGATATGAGCCGACACAGACGAGAACAGAGCTGGCGCT	1286	
DB	1490	GCCCGTACCTTTGTGACCAATTCACATATGAAAG-----CAGACCTGGGGGC	1537	
QY	1287	GAAATCTTTCAGATCTGTGTGTAACAGATACCTAGCTGATCATCTACAGGAG	1346	
DB	1538	GCGGATATGAGAGCTGATCTGTGTCAACAGTTACTCTGTGTATCAGCATATGGGA	1597	
QY	1347	GATTAGAGAGAGCTTCAAGCTCTGAAAAAGACATGAAAGCTTATGCAAGGACAG	1406	
DB	1598	CATGGAGTGTATCAGCTGACCTGGCCACGAGTTTGAACCTGGTATAGACCTACCA	1657	
QY	1407	AAAGCCATCTTGTTCACAGAAATTCGTTGCGGACGAGAGTGGATCCATAGATTC	1466	
DB	1658	GAAACCAATTAATTCAGAGAGGAGTACGGGCGAGAGAAATTTGAGGCTTCCACAGATCC	1717	
QY	1467	ACCTCAATATGTTCTCCGAGAGATACCAAGCAGAGCTGTTGAAGACATCAGGCTCT	1526	
DB	1718	ACCTTGATGTTCAAGTGAAGATACCAAGAAAGTCTGCTGAGAGATATCATTTGCT	1777	
QY	1527	TTTGA-----AAAAGCTATCATTCGAAACACGCTGGGCTTTGAGATTTTA	1580	
DB	1778	GGATCAAAAGCCAAAGAAATATGTGTTGGAAGCTCATCTGAAATTTTGTGATTTAT	1837	
QY	1581	GACTCTCGAGATGTGAAGAACCATTTCTCAACCAAGAGGTGTTTTCACAAAGACAG	1640	
DB	1838	GACTGACAGTCAACACAGAGAGCAGTAGAGGAAGAAAGGACATTTCACTGCGCAGAG	1897	
QY	1641	ACAAACCAAACTGTTGCTCATGTACTGAGA 1671		
DB	1898	ACAACCAAAAGGGGCGCTTCTTTTGGGA 1928		

RESULT 6
AF084552
LOCUS AF084552 2155 bp mRNA linear PRI 07-SEP-1998

DEFINITION	Chlorocebus aethiops beta-glucuronidase mRNA, partial cds.		
ACCESSION	AF084552		
VERSION	AF084552.1 GI:3549608		
KEYWORDS			
SOURCE	Cercopithecus aethiops (African green monkey)		
ORGANISM	Cercopithecus aethiops Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Cercopithecus.		
REFERENCE	1 (bases 1 to 2155) Vervoort R. Partial cDNA sequence of Cercopithecus aethiops (COS7 cell) beta-glucuronidase		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 2155) Vervoort R. Direct Submission		
JOURNAL	Submitted (16-AUG-1998) Cell Genetics, Medical Research Council Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, U.K.		
FEATURES	location/Qualifiers		
source	1..2155 /organism="Cercopithecus aethiops" /mol_type="mRNA" /db_xref="taxon:9534" /cell_line="COS7" 1..1948 /EC_number="3.2.1.31" /function="hydrolase" /note="similar to Homo sapiens beta-glucuronidase; beta-D-glucuronoside; glucuronosylhydrolase; glucuronidase glucuronohydrolase" /codon_start=2 /product="beta-glucuronidase" /protein_id="AAC34593.1" /db_xref="GI:3549609" /translation="GLAAMAVLGPILWGICALALOGMLYPRSSQSRKRLDGLMS RADPSNRGRPEQWVRPLRESGPTLDMVPSFNDISODMLRHRVGVVERE ILPERWTDLSRVLRTGSAHAATVAVNGVHTLREHGTLPREDLSNVQVGLP SHVRLTAINNTLSTLTPGTIQLDLSIKYPKGYFIONTPDFENYAGLQSVLLY TPTPAVIDITVTGVEHDTGLVNYQISVKOSNLELEVRLLDLENKLVANGTIOGL LKVPGARLMPVYLMEHREPAYLSLEVRLLAQTSLSGPDFTLLPVGIRTVAVTESQFL INGKPFYPHGNKHEADIRGKGFMDPLVDFNLRLVLANAPTSHTPYABEVLQ CDRYGIVVIDECPGVGLALPQFNPNVLSNMRVMEVRDKHNPAYMSVANEPA SHLSAGYLLKMTVITHKALDPSRPVFTNNTSNAADGAYVDVTCINSTYSRHDY GHLELQROLTTQFENMYKTYOKPLIQSEYAEATVGHQDPPLMFEETSEYKSLSEY HVVLDQKRYVVELIMNFPADFMTEQSPFVLANKGVFTROQPKSAATLRRRY KIANETRRPHSIAQSCLNSPPT"		
ORIGIN	polyA_signal polyA_signal 2103..2108 2126..2131		
Query Match	10.3%; Score 174.4; DB 9; Length 2155;		
Best Local Similarity	51.6%; Pred. No. 1.5e-35;		
Matches 487;	Conservative 0; Mismatches 436; Indels 21; Gaps 3;		
QY	737	TGGAATTTGAAAAAGCAGATGACCTTGACATCGAATCAAGATCAGCTGGAGC	796
DB	930	TGGGCTGTGTCTCACTTACACCTCCCTGTGGGATCCGACCTGTGCTCAACG	989
QY	797	AGAAGAGCTCATCTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGGAG	856
DB	990	AAAGCAATTTCTCTATCATAGGAAACCTTTCTATTTCCAGGTCTCAACAGCATGGG	1049
QY	857	AATTCCTCCGTTCTGGGAGGAGGACCTTTTATTCATTGATGATTAAGACTTCACTTC	916
DB	1050	ATGCGCATCTCGAGGAGGAGGCTTGCAGCTGGCGCTGTGTGAAGACTTCAACCTGC	1109
QY	917	TGAATGATCAACGCGAATTTTTCAGGACCTCTCATATCTTACAGTGAAGATGCC	976
DB	1110	TTGCTGCTGTGGCCATCTCCGACACAGCATACCTCTACCCGACGAGGAAGTGC	1169
QY	977	TGGAATCTGGCAGACGACTCGGAATCTTGTGATGAGAGAGCCCGGACGTTGTATCA	1036

Db 1170 TCAGATGHTGACCGCTATGGGATTGTGTCATGCATGATGATGCTCGGCGTGGCCTGG 1229
Qy 1037 CAAGGTACCA---CTAACATCCCGAGACTCAGAAAGATAGCAGAAACAACATTAAGAA 1093
Db 1230 CACTCCCGGAGTTCTTCAACAAAGTCCCTGCAGAAACCAATGCGGGTGAAGAGAG 1289
Qy 1094 TGATGCAGACAGACAGAAACCATCCAGTGTATCATGTGATGATGCGGAGCAACAG 1153
Db 1290 TGGTGGCAGGAGCAAGAACCAACCCGCGCTGTATGTGTGTGTGGCAACGAGCTGG 1349
Qy 1154 AGTCCACCATCCAGACCGCGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAA 1213
Db 1350 CGTCCACCTTAAGATCTGCGGCTACTACTTGAAGATGTGTATCATCTCAACCAAGCCT 1409
Qy 1214 TGGATGGAACAGCCGCCCTGTGTATGTGTAGATGATGAGCCCAACAGCAGAAACA 1273
Db 1410 TGGACCCCTCCGCGCTGTGACCTTGTGACCAACTCCAACTATGACG----- 1457
Qy 1274 GAGAGTGGCGCTGAACTTCTGACATGCTGTGTGAAACAGTACTACGGCTGTACA 1333
Db 1458 CAGACAGAGGGGCTCCGTAATGTGACGTATCTGTTGAACAGCTACTCTTGTGATC 1517
Qy 1334 TCTATCAGGGAAGATAGGAAGAGACTTCAAGCTCTGMAAAAGACATAGAGAGCTCT 1393
Db 1518 ACGACTATGGGACCTGAGGTGATTCAAGCGGACAGTTACACAGTTTGAGAACTGGT 1577
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTGTGCACAGAAATTCGTCGGACCGGATAGCTGCA 1453
Db 1578 ATTAAGACGTATCAGAAAGCCCATTTATTCAGAGCGAGATAGAGCGGAAACGATTTGGGT 1637
Qy 1454 TCCACTAGCATCCACTCAATATGTTCTCGAAGAGTACCAACAGACAGCTGTTGAAAA 1513
Db 1638 TTCAACAGGACCGGCTCTGATGTTCTACAGAAAGTACCAAGAGTCTGCTAGAGCAGT 1697
Qy 1514 CGATCAGCTCTTTGAAAAA-----AGACTACATCATCGAAACACAGTGTGGGCT 1567
Db 1698 ACCATGTGTTCTGATCAAAAAACGAGAAAGTACGTGTTGAGAGCTCATCTGGAAT 1757
Qy 1568 TTGCAGATTTTAAAGCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGAGGTGTT 1627
Db 1758 TTGCCGATTTATGACTGAACAGTACCGACGAGAGTGTGGGAAATMAAAAGGGGCT 1817
Qy 1628 TCACAAGACAGACAAACCAACTGTTGCTCATGTAAGTGA 1671
Db 1818 TCACTCGGCAGAGACAAACCAAAAGTGCAGCGTCTCTTTGGCA 1861

RESULT 7
CQ575703 2128 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 3461 from Patent WO0171042.
DEFINITION CQ575703
ACCESSION CQ575703
VERSION CQ575703.1 GI:41639115
KEYWORDS
SOURCE
ORGANISM Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydriidea; Drosophilidae; Drosophila.
REFERENCE
1 Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
Detection kites, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
Patent: WO 0171042-A 3461 27-SEP-2001.
PE Corporation (NY) (US)
Location/Qualifiers
1..2128
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Query Match 10.3%; Score 174; DB 6; Length 2128;

Best Local Similarity 51.6%; Pred. No.1.9e-35;
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;
Qy 734 AGTGGAACTTGAAAAAGCAGATACACTCTGGACATCGGAATCAGAACATCAGCTGG 793
Db 1104 AGTTGAATCTCTCAGAGATACCTACCGCCCTCCAGTGGGATCGTAGCTTAAGTTGG 1163
Qy 794 ACGAAGAGGCTCTATCTGAAACGGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAG 853
Db 1164 ACAACGACAGCTCTGCTGTGATGAAAAACCCCTCTATCTCGGGGATTTGACGACG 1223
Qy 854 AGGAATTCCTCCGTTCTGGGGCAGGCACTTTTATCCATGTGATGATTAAGACTTCAAC 913
Db 1224 AGAATCCGATATCCGGGAAAGGATTTGATTAATGCGCTTCTGTAGATTTTAACC 1283
Qy 914 TTCTGAAGTGAATCAAGCGCAATCTTTAGAGACTCTCATATCTTCAAGTGAAGT 973
Db 1284 TGCTGAAGTGAATGAGCCAAATCATATGCACTCTCATCTATCTTATTCGAAAGT 1343
Qy 974 GGCTGATCTTGCCGACAGACTCGGAATCTTGTGATAGAGAGCCCGCAGCTGGTA 1033
Db 1344 CAATGCACTTGGCCGATCAGCATGTATGATTAATTTGACGAATGCC-----TGCTG 1397
Qy 1034 TCAGAGGTACCACTTAATCCGAGACTCAGAAATAGCAGAAAGACAAATTAAGAA 1093
Db 1398 TCAATATAGATATCTTGAGCCGAGCTACTGGAATATCATATGCTCTCGTGGAGCAAC 1457
Qy 1094 TGATGCAGACACAGAAACCATCTCCAGTGTATCATGTGAGTGTGGCAACGACAG 1153
Db 1458 TGATTCACCGGAGACAGAAACCATCCAAAGTGTGTTGAGTGTGCTGATGCGCAAGCGCA 1517
Qy 1154 AGTCAACCATCCAGACCGCGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAA 1213
Db 1518 GATGAAACAGCAGAGAGCCCTTAATATCTTGAATTCCTGTAACTATGTAAAGAGAA 1577
Qy 1214 TGATGCAGACCGCCCGTTGTCTATGTGATGATGAGACGACAGACAGAGAAACA 1273
Db 1578 TAGCTCAGGACGACCTTAAACCGCGCTAATAATGCAACTC-----TTCCA 1625
Qy 1274 GAGAGTGGCGCTGAAGTACTTTCAGCATCTGTCTGTGAAACAGTATCTACGCTGTACA 1333
Db 1626 GTTCCATTTGGCCGAGTTCTTGACATCTGGGTTTCATTCGCTATTAACCTTTGTGATC 1685
Qy 1334 TCTATCAGGGAAGATGAAGAAAGACTTCAAGCTCTGMAAAAGACATAGAGAGCTCT 1393
Db 1686 AGAATCGGCGCTATGACATGATTAATCTTTGACATGAGAGCCAGAGTTGGC 1745
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTGTGCACAGATTCGATGCGACGCGATAGCTGCA 1453
Db 1746 GGGATTAAGTTGGAAGCGCTGTCAATCAATTTGAGTACGGCGGACACATATGAGGGCA 1805
Qy 1454 TCCACTAGATCCACCTCAATGTTCTCCGAAGTACCAAGCAGAGCTCT----- 1505
Db 1806 TGCACTCACTTCCGCTTTATTTGTGTGGAGAAATCCAGATTGAGCTCTTCTCGCC 1865
Qy 1506 -TGAAGAAGCATCAGGCTCTTTTGAAGAAAGACTCATCATCGSAAACACAGTGTGG 1564
Db 1866 ATTTCAAGCTTTTGACAGAGTACGCGGAAGAGATGTTTATTTGAGAGATTTGTTGA 1925
Qy 1565 CTTTGCAGATTTTAACTCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGAGGTG 1624
Db 1926 ACTTCGCGGATTTTTCGAGCGCGACACTATTAACCGGTGGGCAAAAGAGAG 1985
Qy 1625 TTTTCAAGAGACAGACAAACCAACTCTGTTCTCATGTACTGAAAGACTGT 1678
Db 1986 TCTTTCAAGAGAACCGAACCAACCAAGAGTGTCTCATATTTTAAAGCGGGGT 2039

RESULT 8
BC014142 2182 bp mRNA linear PRI 29-JUN-2004
LOCUS Homo sapiens glucuronidase, beta, mRNA (cDNA clone MGC:20659
DEFINITION IMAGE:466201), complete cds.

ACCESSION	BC014142	GI:40225978
VERSION	BC014142.2	
KEYWORDS	MGC.	
SOURCE	Homo sapiens (human)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 2182)	
AUTHORS	Klauser,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Butow,K.H., Schaefer,C.F., Bat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marinina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stopleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheer,T.E., Brownstein,M.J., Ueda,T.B., Toshiyuki,S., Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Wotley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulya,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Heltón,E., Kettelman,M., Madan,A.C., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Scherchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalka,U., Smolins,D.E., Schercher,A., Schein,J.E., Jones,S.J., and Marra,M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
PUBMED	12477932	
REFERENCE	2 (bases 1 to 2182)	
AUTHORS	Strauberg,R.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT	On Dec 19, 2003 this sequence version replaced gi:15559559. Contact: MGC help desk Email: gcgdb-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Ahter,N., Ayle,K., Beckerson-Steinberg,S.M., Benjamin,B., Blakeley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooke,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karling,R., Kwong,P., Latic,F., Legaspi,R., Maduro,Q.L., Maestello,C., Maestri,B., Maestrian,S.D., McCloskey,J.C., McQuell,J., Pearson,R., Stantipod,S., Thomas,P.J., Touchman,J.W., Targerson,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Series: IRAL Plate: 30 Row: 0 Column: 15 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504222.	
SOURCE	Location/Qualifiers 1..2182 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:20659 IMAGE:4662011" /tissue_type="Colon, adenocarcinoma" /clone_lib="NIH MGC_15" /lab_host="DH10B-R" /note="vector: pOTB7"	

ORIGIN

gene	1. 2182
	/gene="GUSB"
	/note="synonym: MPS7"
	/db_xref="LOCusID:2390"
	/db_xref="MIM:253220"
	2. 1957
	/gene="GUSB"
	/codon_start=1
	/product="glucuronidase, beta"
	/protein_id="AAH14142.1"
	/db_xref="GI:15559560"
	/db_xref="LOCusID:2390"
	/db_xref="MIM:253220"
	/translation="MARGSAVMAAALGPELLMGALGIQGMLYPOESPSPRECKELDGLI WSFPAEDSDNRRRGFEQWYRRPLMESGPVDMPEVPSFNDISQDMRLHFVWVWYVE REVILPERMOTDLRTVRLVIGSHASVAILVWNGVDLLEHGGLPREADISNLVIGVG PLERLRLTIAINNTLPTPLPPTLOYLTDIATSPKGYFVONTYPPFPYAGLGRSIV LILYTPPTIYIDILVTTSVEODSGLVNYQISVKSNI.PKLEVRLLDENKVVANGTGT QGLQLKPVGSLMPYLMHERPAVLYSLEVLDTAOTSIGPVSDFTLLPVGIRTVAVTS QGLFNGRPLYPHNGKIEDADIRKGDMELVLTLMFLWLGNAETRYIPPAEEV MOMDERAGIVIDECPEVGLALPOFPENNVS.LHHHMFMEVVRDKNHPRVMMVSVAN EASHESAGYLLKMLVAHTRKSLDPSRVFVSNVAAKGAAYUVICLNSYSVYV HDHCHLEIOLATOLEPNNYKKYOKRDI0SEVGAETIAGFPHDPRPLMFEYKSLIL BOYHIGLDQKRRKTVSEVILMWFADTEFOSPTVILANKKGI.FTRQRPXSAAPLNE RYWKLINETRYPSHVAKSQCLENSFLT"

ORIGIN	Query Match	10.3%	Score 174;	DB 9;	Length 2182;
Best Local Similarity	52.2%	Pred. No. 1.9e-35;			
Matches 493;	Conservative 0;	Mismatches 430;	Indels 21;	Gaps 4;	
QY	737 TGAACCTTGAAMAAAGACGATACACTCTGGACATCGGAATCAGAACGATCAGCTGGAGC	796			
Db	939 TGGGCGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGGCTGTCAACA	998			
QY	797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGAAAGACAGAG	856			
Db	999 AGACCCAGTTCCTCATCAATGGGAACCTTTCTATTCCACGGGTCAACAGCATGAGG	1058			
QY	857 AATTCGCCGTTTGGGCGAGCGACCTTTATCATGTATGATTAAGACTTCAACCTTC	916			
Db	1059 ATGCGGACATCCGAGGGAAGGGCTTGCATCGCGCTGTGGTGAAGACTTCAACCTCG	1118			
QY	917 TGAAGTGATCAACGCAATCTTCTTCAAGACCTCTCACTACTTCTTCAAGTGAAGATGGC	976			
Db	1119 TTGCGTGGCTTTGGGCCAACGCTTCCGTACAGCACCTACCTTATGCAGAGGAATGA	1178			
QY	977 TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGCAAGCCCGCAGCTTGGTATCA	1036			
Db	1179 TGCAGATGTGTGACCGCTATGGGATGTGTGCATGCATGATGTGTCGCCGGTGGCGCTGG	1238			
QY	1037 CAAGGTACCA---CTACATATCCGAGACTCAGAAAGATAGCAAGAACCACTAAGAAGAA	1093			
Db	1239 CGCTGCCGAGTCTTCAACACGTTTCTGTGATCAACACATGCAAGTATGGAAAG	1298			
QY	1094 TGATCGACAGACCAACCAACCATCCAGTGTGATCATGTGAGATGGGCGAACGAACTGAG	1153			
Db	1299 TGTGTGCTAGGAGACAAAGACCAACCCCGCGGTCTGATGTGTGTCTGTGGCCACAGGCTTG	1358			
QY	1154 AGTCCAAACATCCAGAGCGCGAGGGTTTCTTCAAGCCCTTTATGAGACTGCGCAATGAA	1213			
Db	1359 CGTCCCACTTAAGATCTGTGGCTACACTTGAAGATGTGATGGCTCACACCAAAATCCT	1418			
QY	1214 TGGATCGAACACGCGCGCTGTCTCATGTGTGACATGATGACGACACGACGAGAGAACAA	1273			
Db	1419 TGGACCCCTCCCGGCTGTGACCTTTGTGTGACCACTTAACTATGCG-----	1466			
QY	1274 GAGACGTGGCGCTGAAGTACTTGGACATCTGCTGTGTGAACAGGTATCTAGCGCTGTACA	1333			
Db	1467 CAGACAAAGGGGCTCCGTATGTGTGATGTATCTGTTTGAACAGCTACTCTTTGGTATC	1526			
QY	1334 TCTATCAGGGAAGGATGAAGAAAGACTTCAAGCTCTGAAAAAAGACTTAAGAGCTCT	1393			

Db 1527 ACGACTAGCGGACCTGAGTGTGATTCAGCTGAGCTGGCCACCCAGTTTGAAGACTGTT 1586
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTACAGAAATTCGTTCGAGCGGATAGCTGGCA 1453
Db 1587 ATAGAAAGATCAGAAAGCCCATTTATTCAGAGGAGATGAGACAGAAACGATTCAGAGGT 1646
Qy 1454 TCCATACGATCCACCTCAATATGTTCTCCGAGAGATACCAACAGAGCTCGTTGAAAAG- 1512
Db 1647 TTCACCAAGATCCACCTCTGATGTTTACCTGAGGAGATCCAGAAAGTCTGCTAGAGCACT 1706
Qy 1513 ACGATCAGGCTCTCTTTGAAAAA---AGACTACATCATCGGAAACACGCTGTGGGCTT 1567
Db 1707 ACCATCTGGGCTCTGATCAAAAAACGAGAAATACGTGGTTGAGAGCTCATTTGGAATT 1766
Qy 1568 TTGCGATTTTAAAGCTCCTCGAATGTGAGAAAGCCATTCCTCAACCAAGAGGTGTTT 1627
Db 1767 TTGCGGATTTCACTGACTGAACAGTCACGACGAGAGTGTCTGGGAAATATAAAAGGGGATCT 1826
Qy 1628 TCACAAGAGACAGACAAACCAACTGCTGCTCATGTAAGTCTGAGA 1671
Db 1827 TCACCTCGGACAGACAAACCAAAAGTCAGCTCTTTGGCA 1870

RESULT 9
BT010073 2321 bp mRNA linear INV 12-AUG-2003
LOCUS Drosophila melanogaster LD10588 full insert cDNA.
ACCESSION BT010073.1 GI:33589549
VERSION BT010073.1
KEYWORDS FLI CDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydriidae; Drosophilidae; Drosophila.
1 (bases 1 to 2321)
Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
Champe R., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise B.,
George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
Miranda A., Mungall C.J., Nunoo U., Pacleb J., Paragae V., Park S.,
Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.
and Ceiniker S.

TITLE Direct Submission
JOURNAL Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
COMMENT Sequence submitted by:
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES
source location/Qualifiers

1..2321
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="Y; cn bw sp"
/db_xref="taxon:7227"
1..2321
/gene="CG2135"

CDS

/db_xref="FLYBASE:FBgn0039874"
64..2124
/gene="CG2135"
/note="Longest ORF"
/codon_start=1
/product="LD10588p"
/protein_id="AA022542.1"
/db_xref="GI:33589550"
/db_xref="FLYBASE:FBgn0039874"
translation="MKSHMPVTLTGSLGLVLAATYENDPFLVANKELTDSPLFEX
DNVAKPTKGLYLPDSEETREVSLDGMQVRSDDYDPLQGIKRMFMDALKTGR
ELISMPVASYNDITDNLRDHGTWYERTFVPRSMKTMRTWLPFSVHYAVW
INGANATSHSIGHLPFSEISGLSFGDNRIIVMCDNRLSNRTIPOGSSVKAIVDQ
TVPLOSYPFENYAGIHRSHVLTTPDLHLISOLEVTTOLTKEGIGIDRVMDASK
EGLOIORIOLRVOLRDGQVAAQOINKAVAKVATLIPNATPMPVPLMSPGYLVNL
QPELPVANSKEKLESIDPTTLPVPGHISLSNDNSLLINKGPLYLRGFRBESDIDG
KGLDNALLARDPILKWTGANAARTSHYPTSESMQPADQGLMIIDECVAVNIDIFES
POLLENMSSLEQLIHRDRNHPVAVSVANPEPSNKGALKEFPLVNVVREIAHGR
PLTPAIVANSSCHLAQFLDIVGFNRVNSWYONGRTDMLINLTI BAQSVRRDRGPR
VIOEYGGDTMEGKHSILPAFIWSEYVVELFSPRHKAPDELGRGWFGEVVMRADP
RTAQTITRVGNKKGVFTRNRQPREVAHILRRYFALAKELDMRGLEBDLTVISKHI
HNEL"

ORIGIN

Query Match 10.3%; Score 174; DB 3; Length 2321;
Best Local Similarity 51.6%; Pred. No. 1.9e-35;
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

Qy 734 AGGTGAACTTGAAAAAGACAGTACCTCTGACATCGGAATCAGAACGATGAGCTGGG 793
Db 1100 AGTTGGAATCTCTGACAGATACCTACCGCCCTCCAGTGGGATAGTGGTTAAGTTGGG 1159
Qy 794 ACGAAGAGGCTCTATCTGAAACGGGAACCTGCTTTTGAAGGCTTTGAAAAACG 853
Db 1160 ACAACGAGGCTCTGCTGATGAAAACCCCTCTATCTGCGGAGATTGAGACGCGCAG 1219
Qy 854 AGGAATTCCTCCGTTCTGGGAGGCGACCTTTATTCATGATGATAAAGACTTCAAC 913
Db 1220 AGGACTCGGATATCCGGGGAAGGATTTGATATGCGCTTCTGTAGAAATTTTAAAC 1279
Qy 914 TTCTGAAGTGTACACGCGAATTTCTTTCAGACCTTCACTATCTTTACAGTAAAGT 973
Db 1280 TGCTGAAGTGTACGTGAGCCATCATATCGCACCTCTCACTATCTTTATCCGAAGAGT 1339
Qy 974 GGCCTGATCTTGGCGACAGACTCGGAATCTTTGATGATGACGAACCCCGGACGTTGTA 1033
Db 1340 CAATGCAAGTTTGGCGATCAGATGATGATTAATTTGAGGAATGCCCC-----TGCTG 1393
Qy 1034 TCACAAGGTACCTATCAATCCGAGACTCAGAAAGATGACGAAGACAAACATAAGAA 1093
Db 1394 TCATATATGATATTTCTTGAGCGCGAGCTACGTGAGAAATCAATATCTCTCGCTGAGCAAC 1453
Qy 1094 TGATGACAGACAGAAAGAACCATCTCCAGTGTGATCATGTGAGAGTGTGCGGAACGACAG 1153
Db 1454 TGATTCACCGGGAGACGAAGAACCATCAAGTGTGTTGATGATGCGTGTGCGTGTGCGG 1513
Qy 1154 AGTCCAAACCATCCAGCGGAGGTTTCTTCAAGCCCTTTATGAGACTCCCAATGAAA 1213
Db 1514 GATGAACCAACAGAGGAGCCCTTAATAATCTTTGAATTTCTGTAAACTATTAAGAGAAA 1573
Qy 1214 TGATTCGAACACGCGCGTTGTCTCATGTGAGCATGATGAGACGACACAGAGAAACAA 1273
Db 1574 TAGCTCACGAGACGACTTAAACCGCGGTATTAATGCAACTC-----TTCCA 1621
Qy 1274 GAGACGTGGCGCTGAAGTACTTTCACATCGTCTGTGTAACAGGTACTACGCTGGTACA 1333
Db 1622 GTTGCCATTGGCGCCAGTTTCTGACATCGTGGGTTTCAATCGCTAATCTTGTGTATC 1681
Qy 1334 TCTATCGGGAAGATGTAAGAAAGACTTCAGACTCTGGAATAAGACATATAAGAGCTCT 1393
Db 1682 AGAAGCTGGGCGGTACTGACATGATTAACCTGTGTAACGATAGAGGCCCAAGATTGGC 1741
Qy 1394 ATGCAAGGACAGAAAGCCATCTTTGTACAGAAATTCGATGCGGACGCGATAGCTGGCA 1453

EPASHLESAGYLLKXVIAHTKSLDPSRPVTEVNSNVAADKAPYVDVICLNSYSXW
HDYGHLELIQLOLATOFENWYKYOXKPIIOSXYGKPTIAGPHODPLMETBEYOKSL
BOYHJGLDOKRKXVYVGLIWMFADPMTEOSTRVLGNKKGIPTHQRPKSAFLIRE
RYMKIANETRYHSHVAKSOCLENSEPT"

ORIGIN

Query Match 10.2%; Score 172.4; DB 6; Length 1956;
Best Local Similarity 52.1%; Pred. No. 5e-35;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```
QY 737 TGGAACTTGAAGAAACGAGTCACTCTGGACATCGGAATCGAAACGATCAGCTGGAGC 796
DB TGGGGCGTGTCTGTCTCTACACTCTCGTGGGGATCCGCACTGGCTGTACCA 997
QY 797 AGAAGAGGCTATCTGAACGGGAAACCTGCTTTTGAAGGGCTTTGSAAGCAAGAG 856
DB AAGAGCAGTTCTCTCATCATGGAACCTTTCTATTTCCAGGTGCAACAAAGCATGAGG 1057
QY 857 AATTCCCGCTTGGGGCAGGGCACCCTTTATCATGATGATPAAAGACTTCAACTTC 916
DB ATGCGGACATCGAAGGGAAGGGCTTCGACTGGCCGTGTGTAAGAGACTTCAACTTC 1117
QY 917 TGAAGTGATCAACGGAAATTTCTTCAGGACTCTCATCTATCTTACAGTGAAGTGGC 976
DB TTGCGCTGGCTGTGTCMAACGCTTTCCTGACCACTACCCCTATGACAGAGAACTGA 1177
QY 977 TGGATCTTGGCAGACCTCGGAATCCTGTGATGAGAGAACCCCGACGTTGATCA 1036
DB TGCAGATGTGTACGCGCTATGGAATTTGTGTATGATGATGATGATGATGATGATGATG 1237
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAACAACATPAGAA 1093
DB CGCTGCGCAGTTCTTCAACACGTTTCTCTGCATACACACATGAGGATGAGAAAG 1297
QY 1238 CGCTGCGCAGTTCTTCAACACGTTTCTCTGCATACACACATGAGGATGAGAAAG 1297
DB 1094 TGAATGACAGACAGAAACCATCCAGTGTGATCATGTGAGTGTGGCGAAACGAC 1153
QY 1238 TGGTCGTAGGAGACAGAAACCAACCCGCGTGTGATGTGTGTGTGGCGAAACGAC 1357
DB 1154 AGTCCAAACCATCCAGACGCGGAGGGTTTCTCAAAAGCCCTTATGAGACTGCAATG 1213
DB 1358 CATTCCCACTTAATATCTGTGCTACTACTTGAAGTGTGATCGCTACACCAATCTT 1417
QY 1214 TGGATCGAACACGCCCCGTTGTCAATGTGAGCATGATGAGCAGCAGCAGAGAAACA 1273
DB 1418 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTCAATGATGAG----- 1465
QY 1274 GAGAGCTGGCGCTGAAGTACTTGCACATGTCTGTGTGAACAGTACTACGCTGTACA 1333
DB 1466 CAGACAGAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTCTTGTG 1525
QY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGGAAAAAGCATAGAAAGCTCT 1393
DB 1526 ACGACTAGGGGACCTGAGTGTGATTCAGCTGACGTGACCCAGCTTTGAACTGCT 1585
QY 1394 ATGCAAGGACAGAAAGCCCATCTTGTTCAGAAATTCGGTGCAGACGGGATAGCTGCA 1453
DB 1586 ATPAAAGATATCAGAAAGCCCATTTATTCAGAGGAGATAGAGCAAGAAAGATGCA 1645
QY 1454 TCCACTAGCATCCACTCAATGTTTCCGAAGATGCCAAGCAGAGCTCGTTGAAAG- 1512
DB 1646 TTCACACAGATCCACTCTGATGTTCATGAAAGATGCAAGAAAGGTCTGCTAGAGC 1705
QY 1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATCATGGGAACACAGTGTGGGCT 1567
DB 1706 ACCATCTGGGCTGTGATCAAAAACGAGAAATATGTGTTGAGAGCTCATTTGAAATT 1765
QY 1568 TTGAGATTTTAAAGCTCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACACAGAGG 1627
DB 1766 TTGCGGATTTATGACTGAACAGTCAACGACGAGTGTGGGGAATPAAAAAGGAGATCT 1825
QY 1628 TCACAGAGACAGACACCAAACTGTTGCTCATGTACTAGA 1671
DB 1826 TCACCTCGGAGAGACAAACCAAAAGTGCAGCTTCTTTGCGA 1869
```

RESULT 12

LOCUS CO847932 2191 bp DNA linear PAT 19-AUG-2004
DEFINITION Sequence 237 from Patent WO2004063362.
ACCESSION CO847932
VERSION CO847932.1 GI:51469520
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 Glover, D., Bell, G., Frenz, L. and Midgley, C.
Cell cycle progression proteins
Patent: WO 2004063362-A 237 29-JUL-2004;
Cyclacel Limited (GB)

FEATURES

source
1..2191
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="M15182"

ORIGIN

Query Match 10.2%; Score 172.4; DB 6; Length 2191;
Best Local Similarity 52.1%; Pred. No. 5e-35;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```
QY 737 TGGAACTTGAAGAAACGAGTCACTCTGACATCGGAATCGAAACGATCAGCTGGAGC 796
DB TGGGGCGTGTCTGTCTCTACACTCTCGTGGGGATCCGCACTGGCTGTACCA 1023
QY 797 AGAAGAGGCTATCTGAACGGGAAACCTGCTTTTGAAGGGCTTTGSAAGCAAGAG 856
DB 1024 AAGAGCAGTTCTCTCATCATGGAACCTTTCTATTTCCAGGTGCAACAAAGCATGAG 1083
QY 857 AATTCCCGCTTGGGGCAGGGCACCCTTTATCATGATGATPAAAGACTTCAACTTC 916
DB 1084 ATGCGGACATCGAAGGGAAGGGCTTCGACTGGCCGTGTGTAAGAGACTTCAACTTC 1143
QY 917 TGAATGACAGACAGAAACCATCCAGTGTGATCATGTGAGTGTGGCGAAACGAC 976
DB 1144 TTGCGTGGCTGTGAGGCAAGCCTTTCGTAACACCACTACCCCTATGACAGAGAACTGA 1203
QY 977 TGAATCTTGGCAGACACTCGGAATCCTGTGATGAGAGAAAGCCCGACGTTGATCA 1036
DB 1204 TGCAGATGTGTGACCGCTATGGAATGTGTGTGATGATGATGATGATGATGATGATG 1263
QY 1097 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAACAACATPAGAA 1093
DB 1264 CGCTGCGCAGTTCTTCAACACGTTTCTCTGCATACACACATGAGGATGAGAAAG 1323
QY 1094 TGAATGACAGACAGAAACCATCCAGTGTGATCATGTGAGTGTGGCGAAACGAC 1153
DB 1324 TGGTCGTAGGAGACAGAAACCAACCCGCGTGTGATGTGTGTGTGGCGAAACGAC 1383
QY 1154 AGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAA 1213
DB 1384 CATTCCCACTTAATATCTGTGCTACTACTTGAAGATGTGATGCTCAACCAAAATCTT 1443
QY 1214 TGGATCGAACAGGCCCCGTTGTCAATGTGAGCATGATGAGCAGCAGCAGAGAAACA 1273
DB 1444 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTCAATGATGAG----- 1491
QY 1274 GAGAGTGGCGCTGAAGTACTTCCAGATCGCTGTGTGAACAGGATACTACGCTGTGACA 1333
DB 1492 CAGACAGAGGGGCTCCGTATGTGATGTGATCTGTTGAAACAGTACTCTGTTGATC 1551
QY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGGAAAAAGCATAGAAAGCTCT 1393
DB 1552 ACGACTCGGCACTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1611
```


ACCESSION AX827491
VERSION AX827491.1 GI:39837679
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 Boers, F., Suter-Dick, L. and Wolf, D.
AUTHORS Methode for the toxicity prediction of a compound
TITLE Patent: EP 1344834-A 225 17-SEP-2003;
JOURNAL F. HOFMANN-LA ROCHE AG (CH)
FEATURES
source location/Qualifiers
1..2472
/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"
ORIGIN
Query Match 9.9%; Score 166.8; DB 6; Length 2472;
Best Local Similarity 51.3%; Pred. No. 1.6e-33;
Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;
QY 751 GACGAGTACACTCTGGACATCGGAATCAGATCAGCTGGAGCAGAGAGCTTAT 810
DB 946 GACTTCTATACCTCCCTGTCGGGATTCGACAGTGGCTGCACAAAGGCAAGTTCTTC 1005
QY 811 CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAAATTCCTCGTTCTG 870
DB 1006 ATAAATGGGAAGCTTTCTACTTCCAGGCGTCACACAGCATGAGATTCAATATCCGA 1065
QY 871 GGGCAGGGACCTTTTATTCATTGATGATAAAGACTTCAACCTTCTGAAGTGAATCAAC 930
DB 1066 GGGAGAGGCTTGAAGTGGCTCTGCTGATTAAGATTCAACCTCCGTTGCTCGGG 1125
QY 931 GCGAATTCCTTCAGGACCTCTCATCTTACATGTAAGAGTGGCTGATTTGCCGAC 990
DB 1126 GCAAAATTCCTTCGACACGACCATATCTTCTCGAGAGAGTACTTCACTGTGTGAC 1185
QY 991 AGACTCGGAATCTTGTGTATGACGAAGCCCGCACGTTGTATCAAGAGTACCA--C 1047
DB 1186 CGATATGGAATTTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245
QY 1048 TACAATCCCGAGACTCAAGAGATGACAGAACAATTAAGAAATGATGACAGACAC 1107
DB 1246 TTTGGCAAGTGTCTCTTCGGACCACTAGAGTATGATGATGATGATGATGATGATGATGAT 1305
QY 1108 AAGAACCATCCCAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167
DB 1306 AAAAATCAACCTCGCGT 1365
QY 1168 GACGCGAGGGTTTCTTCAAAAGCCCTTTATGAGACTGCGCAATGAATGATGATGATGATGAT 1227
DB 1366 CTGCGCGGATATTAATTCAAGACGTTGATCGCCACACCAAGCCCTGAGCCCACTCGT 1425
QY 1228 CCCGTTGTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
DB 1426 CCTGTGACCTTTGTGACAAATA-----CCAGATATGACGACAGATGATGATGATGATGATGAT 1473
QY 1288 AAGTATCTTCGATCTGTGTGTGAAGTACTACGCTGTGTATCTATTCAGGGAAG 1347
DB 1474 CCGTACGTGACGT 1533
QY 1348 ATAGAAGAAGACTTCAAGCTTGGAAAAAGCATAGAAGAGCTTATGCAAGGACAGA 1407
DB 1534 CTGAGAGTGTATTCAGCTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGT 1593
QY 1408 AAGCCCATCTTTGTCAAGAAATTCGTTGCGAGCGCGATAGCTGACATCACTACGATCCA 1467
DB 1594 AAGCCCATTTATTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1653
QY 1468 CCTCAATGTTCTCGAAGAGATCAACAGCAGAGCTGTTGAAAAAGCATGATCAGGCTCTT 1527

DB 1654 CTTGCAATGTTCAAGTGAAGATCAACAGAGCTCTCTCGAGAAATTAATTCATTTGATTCCTG 1713
QY 1528 TTGA-----AAAGACTATCATTCGAAACACAGTGTGGCCCTTGGACATTTTAAAG 1581
DB 1714 GATGAAGAAAGAAATATGTCTCGAGAGCTCATCTGGAATTTTGCTGACTTCAATG 1773
QY 1582 ACTCTCAATGTGAAGACCAATTTCAACCAAGGTTTTCACAAAGACAGAGA 1641
DB 1774 ACGAACCAAGTCAACCTGAGATTAACAGAAACAAAGAGGAGATCTTCACTGACAGAGA 1833
QY 1642 CAACCAACTGCTGCTCATGTAAGTGA 1671
DB 1834 AACCCAAAGATGGACCTTCAATTTTGA 1863

Search completed: September 2, 2005, 00:30:40
Job time : 7498 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 19:07:51; Search time 975 Seconds

(without alignments) 10254.813 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atgctaaagaccgcaacgaaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY_NUC Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0% Listing first 45 summaries

Database: N_Geneseq.16Dec04.*

1: geneeqn1980a.*
2: geneeqn1990a.*
3: geneeqn2000a.*
4: geneeqn2001a.*
5: geneeqn2001b.*
6: geneeqn2002a.*
7: geneeqn2002b.*
8: geneeqn2003a.*
9: geneeqn2003b.*
10: geneeqn2003c.*
11: geneeqn2003d.*
12: geneeqn2004a.*
13: geneeqn2004b.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	3	AAA07937 Thermotoga
2	1688	99.9	1692	3	ADS48252 Bacteroid
3	174	10.3	2050	13	ACN43424 Human dia
4	174	10.3	2082	13	ACN43423 Human dia
5	174	10.3	2128	4	ABL04147 Drosophila
6	174	10.3	2169	8	AAD50922 L. mexicana
7	174	10.3	2169	12	ADF47502 Human bet
8	174	10.3	2169	12	ADJ58608 Fusion DN
9	172.4	10.2	1956	4	AAD06386 Human bet
10	172.4	10.2	2191	6	ABK83932 Human CDN
11	172.4	10.2	2191	10	ADD29602 Human tum
12	172.4	10.2	2191	10	ADG89395 Cancer de
13	172.4	10.2	2191	12	ACF57506 Human bet
14	172.4	10.2	2191	12	ADJ62864 Human bet
15	172.4	10.2	2191	13	ADP10326 Reference
16	172.4	10.2	2191	13	ADQ88807 Antagonis
17	172.4	10.2	3525	4	AAE57574 Mutant he
18	172.4	10.2	7328	4	AAE57573 Heavy cha
19	170.8	10.1	3314	2	AAQ58896 Humanised
20	166.8	9.9	2472	10	ADB58226 Toxicity-

21	166.8	9.9	2472	10	ADB52752 Primary r
22	159.6	9.4	1947	4	AAD06387 Murine be
23	145.8	8.6	1929	13	ACN43422 Human dia
24	145.8	8.6	2109	10	ADG76418 Human inc
25	142.2	8.4	23078	4	AAE59508 Propionib
26	142.2	8.4	23078	8	ACF64437 Propionib
27	141.4	8.4	1887	2	AAE59508 Propionib
28	133	7.9	1820	3	AAE59508 Propionib
29	132.6	7.9	1812	3	AAE59508 Propionib
30	130.4	7.7	1888	3	AAE59508 Propionib
31	114	6.7	1814	10	ADG76418 Human inc
32	114	6.7	3451	8	ABK16609 Artificial
33	114	6.7	3451	10	ACC44711 Plasmid p
34	114	6.7	14627	8	ABK16610 Artificial
35	114	6.7	14627	10	ACC44712 Plasmid p
36	110.8	6.6	1809	6	ABK87080 uida DNA
37	110.8	6.6	1812	6	AAH22503 PCR gene
38	110.8	6.6	1812	4	AAH22503 PCR gene
39	110.8	6.6	1812	6	ABK16610 Artificial
40	110.8	6.6	1812	8	ABK16606 Artificial
41	110.8	6.6	1812	10	ACC44709 E. coli b
42	110.8	6.6	1812	13	ADS46099 Beta-gluc
43	110.8	6.6	2001	4	AAH22503 PCR gene
44	110.8	6.6	2001	6	ABK15667 Glucathio
45	110.8	6.6	2001	13	ADR16848 Arabidops

ALIGNMENTS

RESULT 1	AAA07937 standard; DNA, 1689 BP.
ID	AAA07937
AC	AAA07937; (first entry)
DT	26-JAN-2001
DB	Thermotoga maritima beta-glucuronidase gene.
KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KW	Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW	transgenic insect; marker; glucuronide detoxification; de.
OS	Thermotoga maritima.
PN	WO200055333-A1.
PD	21-SEP-2000.
PF	16-MAR-2000; 2000WO-US007107.
PR	17-MAR-1999; 99US-00270957.
PA	(CAMB-) CAMBIA BIOSYSTEMS LLC.
PI	Jefferson RA, Mayer JE;
DR	WPI, 2000-647075/62.
XX	P-PSDB; AAB28406.
PT	Novel microbial beta-glucuronidase genes and gene products used as
PT	reporter/effector molecule, as diagnostic tool, in positive selection, to
PT	target molecules to specific cells and to detect and track linked genes.
PS	Claim 2; Fig 4; 116pp; English.
CC	The present sequence encodes a microbial beta-glucuronidase (GUS)
CC	protein. GUS genes were obtained from six different genera:
CC	Enterobacter/Salmonella, Pseudomonas, Staphylococcus and
CC	Thermotoga. Microbial GUS can be used as a reporter/effector molecule for
CC	transgenic constructions and in in vitro diagnostic applications. It may
CC	also be used to generate sentinel plants that serve as bioindicators of

CC environmental status. It may be used to generate transgenic insects for
 CC tracking insect populations or to facilitate the development of a
 CC bioassay for compounds that affect molecules critical for insect
 CC development (e.g. juvenile hormone). Secreted GUS may also serve as a
 CC marker for beneficial fungi destined for release into the environment. In
 CC animal systems, secreted GUS may be used to achieve extracellular
 CC detoxification of glucuronides (e.g. toxin glucuronide) and to examine
 CC conjugation patterns of glucuronides. Microbial GUS may also be used in
 CC traditional medical diagnostic assays, for drug testing, pharmacokinetic
 CC studies, bioavailability studies, diagnosis of diseases and syndromes,
 CC following progression of disease or its response to therapy. Microbial
 CC GUS has increased thermal stability, high turnover number and enzymatic
 CC activity. It is highly specific for the substrate and water soluble, and
 CC the substrates are stable
 XX

Sequence 1689 BP, 526 A, 374 C, 430 G, 358 T, 0 U, 1 Other;

Query Match 99.9%; Score 1688; DB 3; Length 1689;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAAGACCCGCAACGAAACAAAGAGATTTTCTTATCTTGAATGAGATTGGAAT 60
 DB 1 ATGCTAAGACCCGCAACGAAACAAAGAGATTTTCTTATCTTGAATGAGATTGGAAT 60
 QY 61 CTGTAAGTAAACGACAAAGACAGACCAATGCGCGTTCTGSAAGCTGGAATGAGAGTAC 120
 DB 61 CTGTAAGTAAACGACAAAGACAGACCAATGCGCGTTCTGSAAGCTGGAATGAGAGTAC 120
 QY 121 CAGGATCTGTGCTACGAGAGAGAGACCTTCACTCAAAAACCACTTCTACCTTCCGAAG 180
 DB 121 CAGGATCTGTGCTACGAGAGAGAGACCTTCACTCAAAAACCACTTCTACCTTCCGAAG 180
 QY 121 CAGGATCTGTGCTACGAGAGAGAGACCTTCACTCAAAAACCACTTCTACCTTCCGAAG 180
 DB 121 CAGGATCTGTGCTACGAGAGAGAGACCTTCACTCAAAAACCACTTCTACCTTCCGAAG 180
 QY 181 NAACTTTTCAAAAAACATCTGACATTTTCTTGTGCGGTGAAACGAGCTGCCAGGTC 240
 DB 181 NAACTTTTCAAAAAACATCTGACATTTTCTTGTGCGGTGAAACGAGCTGCCAGGTC 240
 QY 241 TTCCTCAACGAGAGAAAGTGGAGAGATCAATTGAATACCTTCCCTTGGAGTAGAT 300
 DB 241 TTCCTCAACGAGAGAAAGTGGAGAGATCAATTGAATACCTTCCCTTGGAGTAGAT 300
 QY 301 GTGACGGGGAAAGTGAATCCGAGAGAAAGCACTCAGGGTGGTTGTGAGAACAGATTG 360
 DB 301 GTGACGGGGAAAGTGAATCCGAGAGAAAGCACTCAGGGTGGTTGTGAGAACAGATTG 360
 QY 361 AAAGTGGAGAGATTTCCCTCGAAAGTTCCAGACAGCGGCACTCAACCGTGGATTTTTT 420
 DB 361 AAAGTGGAGAGATTTCCCTCGAAAGTTCCAGACAGCGGCACTCAACCGTGGATTTTTT 420
 QY 421 GGAAGTTTTCACTGCAAACTTTCGACTTCCCTACGAGTAGATCAATAAGGCTGTT 480
 DB 421 GGAAGTTTTCACTGCAAACTTTCGACTTCCCTACGAGTAGATCAATAAGGCTGTT 480
 QY 481 CTGATAGATTCAACAGACCGAGAGATTAATCGACATCTGGTGGACACGAGTAGTCT 540
 DB 481 CTGATAGATTCAACAGACCGAGAGATTAATCGACATCTGGTGGACACGAGTAGTCT 540
 QY 541 GAAACGGAGAGAACTTGGAAAAAGTGAAGTAGAAGTCTCAGAAAACCGGGT 600
 DB 541 GAAACGGAGAGAACTTGGAAAAAGTGAAGTAGAAGTCTCAGAAAACCGGGT 600
 QY 601 GGAACGAGAGATGACATCAAACTTGGAGAGAGAGAAAAAGATTAGAACATCAACAGA 660
 DB 601 GGAACGAGAGATGACATCAAACTTGGAGAGAGAGAAAAAGATTAGAACATCAACAGA 660
 QY 661 TTGCTGGAAGGGGAGTTCACTCTCGAAAAAGCCAGGTTCTGAGGCTCGAAGATCCATAT 720
 DB 661 TTGCTGGAAGGGGAGTTCACTCTCGAAAAAGCCAGGTTCTGAGGCTCGAAGATCCATAT 720
 QY 721 CTTTATCTCTCAAGGTGGAATTTGAAAAAGAGAGTAACTCTGGAATCGGAATCAGA 780
 DB 721 CTTTATCTCTCAAGGTGGAATTTGAAAAAGAGAGTAACTCTGGAATCGGAATCAGA 780

QY 781 ACGATCAGCTGGGAGCAGAAAGAGGCTATCTGAACGGGAAACCTGCTTTTGAAGGAC 840
 DB 781 ACGATCAGCTGGGAGCAGAAAGAGGCTATCTGAACGGGAAACCTGCTTTTGAAGGAC 840
 QY 841 TTTGGAAGCAGAGAAATTCCTGTTCTGGGGACGGGACCTTTTATCAATGATGATA 900
 DB 841 TTTGGAAGCAGAGAAATTCCTGTTCTGGGGACGGGACCTTTTATCAATGATGATA 900
 QY 901 AAAGATTTCACTTTGGAATGATCAACCGGAATTTTTCAGAGACCTTCACATATCT 960
 DB 901 AAAGATTTCACTTTGGAATGATCAACCGGAATTTTTCAGAGACCTTCACATATCT 960
 QY 961 TACAGTGAAGTGGAGTGGATCTTGGCGACAGACTCGGAATCTTGTATGACGAAGCC 1020
 DB 961 TACAGTGAAGTGGAGTGGATCTTGGCGACAGACTCGGAATCTTGTATGACGAAGCC 1020
 QY 1021 CCGACGTTGGTATCACAAAGTACCACTCAATCCGAGACTCAGAAAGATGACGAAGAC 1080
 DB 1021 CCGACGTTGGTATCACAAAGTACCACTCAATCCGAGACTCAGAAAGATGACGAAGAC 1080
 QY 1081 AACATTAAGAAAGATGATGACAGACAAAGAACATCCAGTGTGATGATGAGAGTGTG 1140
 DB 1081 AACATTAAGAAAGATGATGACAGACAAAGAACATCCAGTGTGATGATGAGAGTGTG 1140
 QY 1141 GCGAAACGAACAGAGTCCAACTCAGACGCGAGAGGTTCTTCAAAAGCCCTTATGAG 1200
 DB 1141 GCGAAACGAACAGAGTCCAACTCAGACGCGAGAGGTTCTTCAAAAGCCCTTATGAG 1200
 QY 1201 ACTGCAATGAAATGATGACAAACGCGCCGTTGTATGTGATGATGATGAGACGACCA 1260
 DB 1201 ACTGCAATGAAATGATGACAAACGCGCCGTTGTATGTGATGATGATGAGACGACCA 1260
 QY 1261 GACGAGAGAAACAAAGAGCTGGCGGTGAAGTATTCGACATGCTGTGTGAACAGTAC 1320
 DB 1261 GACGAGAGAAACAAAGAGCTGGCGGTGAAGTATTCGACATGCTGTGTGAACAGTAC 1320
 QY 1321 TACGCTGTGATCATCTATCAGGAGAGATAGAAGAGACTTCAAGCTCTGGAATAAAGAC 1380
 DB 1321 TACGCTGTGATCATCTATCAGGAGAGATAGAAGAGACTTCAAGCTCTGGAATAAAGAC 1380
 QY 1381 ATGAGAGGCTCTATGCAAGGACAGAAAGCCATCTTGTACAGAAATTCGGTCCGAC 1440
 DB 1381 ATGAGAGGCTCTATGCAAGGACAGAAAGCCATCTTGTACAGAAATTCGGTCCGAC 1440
 QY 1441 GCGATAGCTGGATCCATCAAGATCCACTCAATGTTCTCCGAAGATACCAAGCAGAG 1500
 DB 1441 GCGATAGCTGGATCCATCAAGATCCACTCAATGTTCTCCGAAGATACCAAGCAGAG 1500
 QY 1501 CTCGTTGAAAGAGATCAGGCTCTTTTGAATAAAGACTACATCATCGAACAACAGCTG 1560
 DB 1501 CTCGTTGAAAGAGATCAGGCTCTTTTGAATAAAGACTACATCATCGAACAACAGCTG 1560
 QY 1561 TGGGCTTTGCAAGATTTTAAAGCTCTCAGATGAGAAACCAATCTCAACCAAG 1620
 DB 1561 TGGGCTTTGCAAGATTTTAAAGCTCTCAGATGAGAAACCAATCTCAACCAAG 1620
 QY 1621 GGTGTTTTCAAGAAGACAGCAACCCAACTCGTGTCTCATGTACTGAGAAAGCTGTGG 1680
 DB 1621 GGTGTTTTCAAGAAGACAGCAACCCAACTCGTGTCTCATGTACTGAGAAAGCTGTGG 1680
 QY 1681 AGTGAAGTT 1689
 DB 1681 AGTGAAGTT 1689

RESULT 2
 ADS48252
 ID ADS48252 standard; cDNA, 1692 BP.
 XX
 AC ADS48252;
 XX
 DT 02-DEC-2004 (first entry)
 XX

DE Bacterial polynucleotide #2995.
 XX
 KM Recombinant DNA construct; transformed plant; improved plant property;
 KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KM pathogen tolerance; pest tolerance; plant disease resistance;
 KM cell cycle pathway modification; plant growth regulator;
 KM homologous recombination; seed oil yield; protein yield; carbohydrate;
 KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KM bacterial polynucleotide; gene; ss.
 XX
 OS Bacteria.
 XX
 PN US200323675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 DR WPI, 2004-061375/06.
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 26682; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plants with
 CC improved plant properties. The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 CC
 SQ Sequence 1692 BP; 527 A; 374 C; 432 G; 359 T; 0 U; 0 Other;
 Query Match 99.9%; Score 1688; DB 13; Length 1692;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATGCTAAGCCGCAACGAAACAAAGAGATTATTTCTTATCTTGAATGAGTTTGAAT 60
 DB 1 ATGGTAAGACCCGCAACGAAACAAAGAGATTATTTCTTATCTTGAATGAGTTTGAAT 60
 QY 61 CTTGAAGTAACGAGAAAGACAGCAATCGCGTTCTTGAAGCTGGAATGAGAGTAC 120
 DB 1141 GCGAAGCAACGAGTCCAAACATCCAGACCGGAGGTTCTTCAAGCCCTTTATGAG 1200

DB 61 CTTGAAGTAACGAGAAAGACAGACCAATCGCGTTCTTGAAGCTGGAATGAGAGTAC 120
 QY 121 CAGGATCTGTGCTACGAGAGAGACCTTACCTTCAAAAACCACTTTACGTTCCGAG 180
 DB 121 CAGGATCTGTGCTACGAGAGAGACCTTACCTTCAAAAACCACTTTACGTTCCGAG 180
 QY 181 NAACCTTTCAGAAAACATCATGACTTTTCTTGTGCGGTGTAACAACGACCTGCGAGTTC 240
 DB 181 GAACCTTTCAGAAAACATCATGACTTTTCTTGTGCGGTGTAACAACGACCTGCGAGTTC 240
 QY 241 TTCCTCAACGAGAGAAAAGTGGAGAGATCATATTGAATCCTTCCCTTCGAAGTAGAT 300
 DB 241 TTCCTCAACGAGAGAAAAGTGGAGAGATCATATTGAATCCTTCCCTTCGAAGTAGAT 300
 QY 301 GTGACGGGGAAGTGAATCCGAGAGAAAGAACTCAAGGTGTTGTTGAGAACAGATTG 360
 DB 301 GTGACGGGGAAGTGAATCCGAGAGAAAGAACTCAAGGTGTTGTTGAGAACAGATTG 360
 QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACACCGGACCTCACACCGTGGATTTT 420
 DB 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACACCGGACCTCACACCGTGGATTTT 420
 QY 421 GGAAGTTTCCACCTGCAAACTTGCATTTCCCTACGCTGAGATCATTAAGCCCTGTT 480
 DB 421 GGAAGTTTCCACCTGCAAACTTGCATTTCCCTACGCTGAGATCATTAAGCCCTGTT 480
 QY 481 CTGATAGAGTTTCAAGACCAACGAGAGATCTCGACATCTGGGTGGAACAGAGTGTCT 540
 DB 481 CTGATAGAGTTTCAAGACCAACGAGAGATCTCGACATCTGGGTGGAACAGAGTGTCT 540
 QY 541 GAACCGGAGAAACCTTGGAAAAGTGAAGTAGAAGTCTTCAGAAACACCGGTG 600
 DB 541 GAACCGGAGAAACCTTGGAAAAGTGAAGTAGAAGTCTTCAGAAACACCGGTG 600
 QY 601 GGAACGAGATGATGATCAATCTTGAAGAGAAAGAAAGATAGAAATCCAAACAGA 660
 DB 601 GGAACGAGATGATGATCAATCTTGAAGAGAAAGAAAGATAGAAATCCAAACAGA 660
 QY 661 TTGCTGGAAGGGAGTTCACTCTGAAAAACGCGAGGTTCTGGAGCTCGAAGATCATAT 720
 DB 661 TTGCTGGAAGGGAGTTCACTCTGAAAAACGCGAGGTTCTGGAGCTCGAAGATCATAT 720
 QY 721 CTTTATCTCTTCAAGTGAACCTTGAAAAACGAGTACATCTTGAACATCGGAATCAGA 780
 DB 721 CTTTATCTCTTCAAGTGAACCTTGAAAAACGAGTACATCTTGAACATCGGAATCAGA 780
 QY 781 ACGATCAGCTGGAGAGAGAGAGGCTTATCTGAACGAGAAACCTGCTTTTGAAGGCT 840
 DB 781 ACGATCAGCTGGAGAGAGAGAGGCTTATCTGAACGAGAAACCTGCTTTTGAAGGCT 840
 QY 841 TTTGGAAGACAGAGAAATTTCCCGTTCTGGGCAAGGCACTTTTATCATTTAGATTA 900
 DB 841 TTTGGAAGACAGAGAAATTTCCCGTTCTGGGCAAGGCACTTTTATCATTTAGATTA 900
 QY 901 AAAGCTTCAACCTTCTGAAGTATCAACGGAATTTCTTCAAGACCTTCTCATATCTCT 960
 DB 901 AAAGCTTCAACCTTCTGAAGTATCAACGGAATTTCTTCAAGACCTTCTCATATCTCT 960
 QY 961 TACAGTGAAGAGTGTGCTGATTTTGCAGACAGACTCGGAATCTTGTGATGAGAGGCC 1020
 DB 961 TACAGTGAAGAGTGTGCTGATTTTGCAGACAGACTCGGAATCTTGTGATGAGAGGCC 1020
 QY 1021 CCGCAGTTGTGATCACAAGGTACCATTAATCCGAGACTCAGAAAGTAGCAGAAAGC 1080
 DB 1021 CCGCAGTTGTGATCACAAGGTACCATTAATCCGAGACTCAGAAAGTAGCAGAAAGC 1080
 QY 1081 AACATTAAGAAATATATGACAGACAAAGAACATCCCAAGTGTATCATGTGAGTGTG 1140
 DB 1081 AACATTAAGAAATATATGACAGACAAAGAACATCCCAAGTGTATCATGTGAGTGTG 1140
 QY 1141 GCGAAGCAACGAGTCCAAACATCCAGACCGGAGGTTCTTCAAGCCCTTTATGAG 1200
 DB 1141 GCGAAGCAACGAGTCCAAACATCCAGACCGGAGGTTCTTCAAGCCCTTTATGAG 1200

QY	1201	ACTGCCAATGAATGGATCTGAACAACGCCCGCTTGATCTGATGTCGATGATGATGAAGCACA	1260
Db	1201	ACTGCCAATGAATGGATCTGAACAACGCCCGCTTGATCTGATGTCGATGATGATGAAGCACA	1260
QY	1261	GACGAGAGAACCAAGAGACGTGGCGCTGAAAGTACTTCGACATCGTCGTGTGAACAGGTAC	1320
Db	1261	GACGAGAGAACCAAGAGACGTGGCGCTGAAAGTACTTCGACATCGTCGTGTGAACAGGTAC	1320
QY	1321	TACGGCTGGTACATCTATCTACGAGGAAGATAGAAAGAGACCTTCAAGCTCTGGAAAAAGAC	1380
Db	1321	TACGGCTGGTACATCTATCTACGAGGAAGATAGAAAGAGACCTTCAAGCTCTGGAAAAAGAC	1380
QY	1361	ATAGAAGAGCTCTTAGTCGAAGGACACAGAAAGCCCATCTTTGTACAGAAATTGCGTGGGAC	1440
Db	1361	ATAGAAGAGCTCTTAGTCGAAGGACACAGAAAGCCCATCTTTGTACAGAAATTGCGTGGGAC	1440
QY	1441	GCGATAGCTGGATCATCTACAGATCACCTCAAAATGTTCTCCGAGAGATGACCAAGCAGAG	1500
Db	1441	GCGATAGCTGGATCATCTACAGATCACCTCAAAATGTTCTCCGAGAGATGACCAAGCAGAG	1500
QY	1501	CTCGTTGAAAAGACGATCAGGCTCTCTTTGAAAAAAGACTACATCATCGGAAACACACGTG	1560
Db	1501	CTCGTTGAAAAGACGATCAGGCTCTCTTTGAAAAAAGACTACATCATCGGAAACACACGTG	1560
QY	1561	TGGGCGCTTTGCAGATTTTAAAGCTCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAG	1620
Db	1561	TGGGCGCTTTGCAGATTTTAAAGCTCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAG	1620
QY	1621	GGTGTGTTTCAACAAGAGACAGACAACCCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG	1680
Db	1621	GGTGTGTTTCAACAAGAGACAGACAACCCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG	1680
QY	1681	AGTGAAGTT 1689	
Db	1681	AGTGAAGTT 1689	
RESULT 3			
ID	ACN43424	standard; cDNA; 2050 BP.	
AC	ACN43424;		
XX	18-NOV-2004	(first entry)	
DT	Human diagnostic and therapeutic polynucleotide SEQ ID NO:2299.		
DE	as; gene; gene therapy; human diagnostic and therapeutic polynucleotide;		
XX	dithp.		
KM	Homo sapiens.		
OS	Homo sapiens.		
XX	WO2004023973-A2.		
PN	25-MAR-2004.		
PD	12-SEP-2003; 2003WO-US028227.		
PF	12-SEP-2002; 2002US-0410259P.		
XX	12-SEP-2002; 2002US-0410260P.		
PR	(INCY-) INCYTE CORP.		
XX			
PI	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;		
PI	Hartshorne TA, Suchorolski MT, Altus CM, Plets SJ, Elder LV;		
PI	Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP;		
PI	Stevens KA, Blanchard JL, Panzer SR, Wang X, Ju AP, Gerstein EH;		
PI	Peralta CH, Anderson SB, Rioux P, Shen ED, Wu MC, Stuve LL;		
PI	Laegre RE, Spiro PA, Stewart EA, Wingrove J, Vilt UA, Kliron ES;		
PI	Xu Y, Kwong M, Policky JU, Hurwitz BL, Ma Y, Jackson JU, Gietzen D;		
PI	Patury S, Shi X, Suarez CJ;		
XX			

DR		WPI, 2004-329368/30.	
XX	P-PSDB; ABM84772.		
PT	New diagnostic and therapeutic polynucleotides and polypeptides, useful		
PT	in diagnosing a condition, disease or disorder associated with human		
PT	molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or		
XX	in gene mapping.		
PS	Claim 1, Page; 190pp; English.		
CC	The invention relates to novel diagnostic and therapeutic polynucleotides		
CC	selected from one of the 2722 sequences defined in the specification. A		
CC	polynucleotide of the invention may have a use in gene therapy. The human		
CC	diagnostic and therapeutic polynucleotides (dihp) or polypeptides may be		
CC	used to diagnose a particular condition, disease or disorder associated		
CC	with human molecules, e.g. cell proliferative disorders,		
CC	autoimmune/inflammatory disorder, developmental disorder, endocrine		
CC	disorder, neurological disorders, gastrointestinal disorders, or		
CC	infections caused by virus, bacteria, fungi or parasite. The dihp		
CC	molecules may also be used in genetic mapping, in identifying individuals		
CC	from minute biological samples, in detecting single nucleotide		
CC	polymorphisms, as molecular weight markers, and for somatic or germline		
CC	gene therapy. The present sequence represents a dihp polynucleotide of		
CC	the invention. Note: The sequence data for this patent is not represented		
CC	in the printed specification, but was obtained in electronic format		
CC	directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm		
XX			
SQ	Sequence 2050 BP; 465 A; 532 C; 598 G; 455 T; 0 U; 0 Other;		
	Query Match 10.3%; Score 174; DB 13; Length 2050;		
	Best Local Similarity 52.2%; Pred. No. 2.7e+3;		
	Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;		
Oy	737 TGGAACTTGAAGGACGATCACTGTGCACATCGAATCAGAAGACGATCCTGGAGC	796	
Db	815 TGGGGCTGTGTTCGACTTTACACACTCCCTGTGGGATCCGACTGTGGCTGCACCA	874	
Oy	797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTGAAGGCTTTGAAGACACAGC	856	
Db	875 AGAGCAATCTCTCATCATATGGGAAACCTTTCTATTTCACGGTGTCACAAGACATGGG	934	
Oy	857 AATTCGCCGTTCTGGGGCAGGGGCACCTTTTATTCATTGATGATTAAGACTTCAACCTTC	916	
Db	935 ATGGCGAATCCAGAGGAAGGGCTTCCACTGGCGGTCTGTGTAAAGACCTTCAACTGC	994	
Oy	917 TGAAATGATCAACGCGAATCTTTCAGGACCTCTCACTATCTTACAGTAAGATGGC	976	
Db	995 TTCGTGGCTTGTGTCMAACGCTTTCGTCACACCACTACCCTATCCAGAGGAATGA	1054	
Oy	977 TGGATCTTCCGACAGACTCGGAATCCCTGTGATGACGAAGCCCCGACGGTGGTATCA	1036	
Db	1055 TGCAGATGTGACCGCTATGGAGATGTGTGATCATGAGATGTCCCGGCGTGGCTGG	1114	
Oy	1037 CAAGGTACCA---CTACATCCCGAGACTCGAAGATAGCAAGAACATAAGAGAA	1093	
Db	1115 CGTCGCGCAGTTCTTTCAAACAAGTTTCTGTGACATCACATGACAGGTGATGGAAGAG	1174	
Oy	1094 TGATGCACAGACACAGAACAATCCCACTGTGATCATGTGAGTGTGGCCAACGACAG	1153	
Db	1175 TGTGTCCGTAGGGAACAAGAACCCCGCGGTGTGATGTGTCTGTGTGCCAACGAGCTGT	1234	
Oy	1154 AGTCCAAACATCCAGAGCGCGAGGTTTTCTTCAAAGCCCTTTATGAGACTGSCAATGAA	1213	
Db	1235 GTTCCCACTTGAATCTGTCTGTGCTATCTTGAAGATGTTATGCTTCAACCAATCTT	1299	
Oy	1214 TGGATCGAACAAGCCCCGTTGTCTCATGTGACATGATGAGACGACACAGACGAGAACAA	1273	
Db	1295 TGGACCCCTCCCGGCGGTGACCTTTTGTGACCACTGAATGAGAG-----	1342	
Oy	1274 GAGAGTGTGGCGCTGAATGATCTTGCACATCGTCTGTGTGAACAAGTACTTAGCGGTGTCA	1333	
Db	1343 CAGACAAAGGGGCTCCGATGTGATGATGATCTGTTTGGAACAGCTACTACTCTTGTGTATC	1402	

QY 1334 TCTATCAGGAGAAAGATAGAGAGACTTCAAGCTCTGGAAAAAGACATAGAGAGCTCT 1393
 DB 1403 ACGACTACGGGACCCCTGAGTGTATTCAGCTGAGCTGGCCACCCAGTTTGAGAGCTGCT 1462
 QY 1394 ATGCAAGCAGCAAGAAAGCCCATCTTTGTCAAGATTCGCTGGAGCGGATAGCTGGCA 1453
 DB 1463 ATTAGAGATATCAGAAAGCCCATATTCAGAGCAGATATGAGAGAGAAAGATTCAGGGT 1522
 QY 1454 TCCATCAGATCCACTCAATGTTCTCGAAGAGTACCAAGAGAGCTCTGTTGAAAAG- 1512
 DB 1523 TTCACCAAGATCCTCTGATGTTTCACTGAAAGATACCAAGAAAGTCTGTAAGAGCAGT 1582
 QY 1513 ACGATCAGGCTCTTTTGAAGAA-AGACTACATCATCGGAACACAGCTGGGGCT 1567
 DB 1583 ACCATCTGGGTCTGATCAAAAACGAGAAATACCTGTTGAGAGCTCATTTGGAAT 1642
 QY 1568 TTGCAAGATTTTAAGACTCTCTCAAGATGTAGAGAGACCCATTTCAACCAAGGCTGTT 1627
 DB 1643 TTGCGGATTTCACTGATGAAAGTCAACGAGAGAGTCTGGGGAATTAAGAGGGATCT 1702
 QY 1628 TCACAAAGAGACAGCAACCCAAAGCTGTTGCTCATGTACTAGGA 1671
 DB 1703 TCACTCGCAGAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1746

RESULT 4
 ACN43423
 ID ACN43423 standard; cDNA; 2082 BP.
 ACN43423;
 DT 18-NOV-2004 (first entry)

Human diagnostic and therapeutic polynucleotide SEQ ID NO:2298.

68; gene; gene therapy; human diagnostic and therapeutic polynucleotide; dthp.

OS Homo sapiens.

PN WO2004023973-A2.

PD 25-MAR-2004.

PF 12-SEP-2003; 2003WO-US028227.

PR 12-SEP-2002; 2002US-0410259P.

PR 12-SEP-2002; 2002US-0410260P.

PA (INCY-) INCYTE CORP.

PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F, Hartsborth TA, Suchorolski MT, Altus CM, Plets SJ, Sider LV, Mooney EM, Deleane AM, Panesar IS, Banville SC, Reddy TP, Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH, Perrella CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL, Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vilt UA, Kirtan ES, Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D, Patury S, Shi X, Suarez CJ;

DR MPI; 2004-329368/30.
 DR P-PsDB; ABM84771.

PT New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.

PS Claim 1; Page; 190pp; English.

CC The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2122 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human

CC diagnostic and therapeutic polynucleotides (dthp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dthp CC molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline CC gene therapy. The present sequence represents a dthp polynucleotide of CC the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format CC directly from WFO at www.wipo.int/pct/en/sequences/listing.htm

Query Match 10.3%; Score 174; DB 13; Length 2082;
 Best Local Similarity 52.2%; Pred. No. 2.7e-43;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGACATCGAATGAAAGATGAGCTGGAG 796
 DB 847 TGGGCTGTGTCTGATCTTCAACACTCCTGTGGGATCCGACTGTGCTGCACCA 906
 QY 797 AGAAGAGCTATCTGAAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGAG 856
 DB 907 AGAGCAATCTCTCATCATGAGGAAACCTTTCTATTTCCAGGTGTCAACAGCATGAGG 966
 QY 857 AATTCCTCCGCTTCTGGGGGAGGAGCACTTTTATTCATGATGATTAAGACCTTCACTTC 916
 DB 967 ATGCGAGATCCGAGGAGAGGCTTCGACTGGCCGCTGCTGTGAGAGACTTCAACCTGC 1026
 QY 917 TGAAGTGATCAAGCGCAATTTCTTTCAGAGCTCTCACTATCTTCAAGTAAAGTGGC 976
 DB 1027 TTGCTGCTGTGTGTCGCAACCTTTCGATACAGCACTACCCCTATGACAGAGATGA 1086
 QY 977 TGAATCTTGGCCGACAGATCTCGGAATCTTGTGATAGAGAGCCCGGACCTTGTATCA 1036
 DB 1087 TGCAGATGTGACCGCTATGAGATGTGTGATCATGAGATGTGCCGCTGGGCTCGG 1146
 QY 1037 CAAGTACCA---CTACAACTCCGAGACTGAGAGATAGCAGAAAGACAACTAAGAGAA 1093
 DB 1147 CGCTGCCGACTTTCTTCAACAACGTTTCTGTGATACACATGACAGGTATGAGAGAG 1206
 QY 1094 TGATGACAGACACAGAAACCATCCAGTGTATCATGTGAGAGTGTGGCAGAACCAAG 1153
 DB 1207 TGTGTCGTAAGGAGCAAGAACCAACCCGCGGTGTGATGTGTGTGGCCAAAGAGCTGG 1266
 QY 1154 AGTCCAACTATCCAGAGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAA 1213
 DB 1267 CGTCCCACTTGAATCTGTGCTGCTACTTGAAGATGTGATGCTTCAACACCAATCT 1326
 QY 1214 TGAATGAAACAGCGCCGCTGTGATGAGATGATGAGAGAGCAGCAGAGAGAGAA 1273
 DB 1327 TGGACCTCTCCGCGCTGTGACTTTTGTAGCACTTAACTATGACG----- 1374
 QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCTGTGTGTGAACAGGTACTACGCTGTGACA 1333
 DB 1375 CAGACAGGGGGCTCCGATGTGATGTGATCTTTTGAACAGTACTACTCTGTTATC 1434
 QY 1334 TCTATCAGGAAAGATGAGAGAGACTTCAAGCTCTGGAAGAAAGCATGAGAGCTCT 1393
 DB 1435 ACGACTACGGGACCTGAGTGTGATTCAGCTGACGCTGGCCACCAAGTTGAGAACTGCT 1494
 QY 1394 ATGCAAGCAGCAAGAAAGCCCATCTTTGTCAAGAAATTCGGGGGAGAGCGATAGCTGCA 1453
 DB 1495 ATTAGAGATATCAGAAAGCCCATTTATTCAGAGCAGATTTGAGCAGAAACGATTCAGAGGT 1554
 QY 1454 TCCACTACGATCACTCAATGTTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAAG- 1512
 DB 1555 TTCAACAGATTCACCTCTGATGTTCACTGAAAGATACAGAAAGTCTGTAAGAGAGT 1614
 QY 1513 ACGATCAGGCTCTTTTGAAGAA-AGACTACATCATCGGAACACAGTGTGGGCT 1567

```
Db      1615 ACCATCTGGGCTCGATCAAAAACGCGAAAATACGTGGTGGAGAGCTCATTTGGAAATT 1674
Qy      1568 TTGCAGATTTTAAAGTCTCTCGAATGTGAGAAAGCCATTCTCAACCAAGGCTTTT 1627
Db      1675 TTGCCGATTTCTGACTGAAAGTCAACGAGAGTGTCTGGGAATTAAGGGGATCT 1734
Qy      1628 TCACAAGAGACAGAACCAACTCGTGTCTCATGTACTGAGA 1671
Db      1735 TACTCTGCGAGAGACCAACCAAAAGTGCAGCGTTCCTTTGGCGA 1778

RESULT 5
ABL04147
ID      ABL04147 standard; cDNA; 2128 BP.
XX
AC      ABL04147;
XX
DT      26-MAR-2002 (first entry)
XX
DE      Drosophila melanogaster expressed polynucleotide SEQ ID NO 6923.
XX
KW      Drosophila; developmental biology; cell signalling; insecticide;
KW      pharmaceutical; gene; ss.
XX
OS      Drosophila melanogaster.
XX
PN      WO200171042-A2.
XX
PD      27-SEP-2001.
XX
PF      23-MAR-2001; 2001WO-US009231.
XX
PR      23-MAR-2000; 2000US-0191637P.
XX
PR      11-JUL-2000; 2000US-00614150.
XX
PA      (PEKE ) PE CORP NY.
XX
PI      Venter JC, Adams M, Li PMD, Myers EW;
XX
DR      WPI, 2001-656860/75.
XX
DR      P-PSDB; ABB60044.
XX
PT      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions.
XX
PS      Claim 1; SEQ ID NO 6923; 21bp + Sequence Listing; English.
XX
CC      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC      ABBS2072). The sequence data for this patent did not form part of the
CC      printed specification, but was obtained in electronic format directly
CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 2128 BP; 546 A; 511 C; 540 G; 531 T; 0 U; 0 Other;
XX

Query Match      10.3%; Score 174; DB 4; Length 2128;
Best Local Similarity 51.6%; Pred. No. 2.8e-43;
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;
```

```
Qy      854 AGGAATTCGCCGTTCTGGGCGAGGCACTTTATTCATGTATGATTAAGACTTCAAC 913
Db      1224 AGGACTCCGATATTCGCCGGAAGGATTTGATTAATGCGCTTCTGTAAGATTTTAAC 1283
Qy      914 TTTCGAAGTGTATCAACGCGAATTCCTTCAGAGCTCTCACTATCTTTACAGTGAAGT 973
Db      1284 TGCTGAAGTGTACAGGAGCAATGATATGCACTCTCACTATCTTTATTCGGAAGT 1343
Qy      974 GCGTGAATCTTGGCGACAGACTCGGAATCTTTGATATGACGAAGCCCCGACGTTGGTA 1033
Db      1344 CAATGCACTTTGGCGATCAGATGATATTAATGATTAATGGAATGCC-----TGCCTG 1397
Qy      1034 TCACAAGGTACCACTACATCCCGAGCTCGAAGATAGAGAAAGACAATTAAGAA 1093
Db      1398 TCAATATGATATATCTTCGAGCGGACGCTACTGGAATACATGCTCTCGCTGAGAC 1457
Qy      1094 TGATCGACAGACAGAAACCATCCAGTGTATCATGTGAGTGTGCGAAGCAACGAG 1153
Db      1458 TGATCCACCGGAGACAGAAACCATCCAGTGTGTGATGATGTCGATCCAGAGCGCGA 1517
Qy      1154 AGTCCACCATCCAGACGCGGAGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db      1518 GATCGAACACACAGGAGCCCTTAATACTTTGAATTCCTGTAACTATGTAAAGAGAAA 1577
Qy      1214 TGATCGAAACAGGCGCGTGTGTGAGACATGATGAGACGACACAGAGAAACAA 1273
Db      1578 TAGCTCAGACGACCTTAAACCGGCTTAATATGCCAATC-----TTCCA 1625
Qy      1274 GAGACGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGGTATACGCTGTACA 1333
Db      1626 GTTGCAATTTGGCGCAGTTTCTGACATCGTGGTTCAATCGGTATTAATCTTGGTATTC 1685
Qy      1334 TCTATCAGGGAAGATAGAAAGACTTCAAGCTTGTGAAAAAGACATAGAAGCTCT 1393
Db      1686 AGAATCTGGGCGGATGATGATGATTAATCTGTGACGATGAGGCCAGAGATTGGC 1745
Qy      1394 ATGCAAGGACAGAAAGCCATCTTTGTGCAGAAATTCGATGCGGAGCGGATAGCTGCA 1453
Db      1746 GGGATAGTTGGGAAGGCTGTCACTCAATTTGATGATGACGCGGAGCACTTATGAGGCA 1805
Qy      1454 TCACATAGATCACTCAATATGTTCTCCGAAGATACCAAGACAGCTGT----- 1505
Db      1806 TGCACTCACTTCCGCGCTTTATTTGTGCGAGGAATACAGATTTGAGTCTTCTCGCC 1865
Qy      1506 -TGAAGAAGCATGAGGCTCTTTTGAAGAAAGACTACATATGGAACACAGTGTGG 1564
Db      1866 ATTTCAGGCTTTGACGAGTTACGCGAGAGATGTTTATTTGAGAGTTTGTGGA 1925
Qy      1565 CTTTGCAGATTTTAAGCTCTCAGATGTGAGAAAGCCATTCTCAACCAAGGCTG 1624
Db      1926 ACTTCGCGGATTTTCGAGCGCGAGACTATTAACCGGTGGGCGCAACAAAAGGAG 1985
Qy      1625 TTTTCACAAGACAGAACCAACCAACTCGTGTCTCATGTACTGAGAAAGCTGT 1678
Db      1986 TCTTTACAAAGAACCGAACCAACCAAGAAAGGCTCACATTTTAAGGGGCGGT 2039

RESULT 6
AAD50922
ID      AAD50922 standard; DNA; 2169 BP.
XX
AC      AAD50922;
XX
DT      02-APR-2003 (first entry)
XX
XX
XX      L. mexicana SAP-human beta-GUS-IGF-II chimeric DNA.
XX
KW      Lysosome; metabolic disease; lysosomal storage disease; gene therapy;
KW      Gaucher's disease; Pompe's disease; Hurler's syndrome; neuroprotective;
KW      Niemann-Pick's disease; Schindler's disease; mucopolysaccharidosis;
KW      Batten's disease; prosoposis; infantile neuronal ceroid lipofuscinosis;
KW      fucosidosis; mannosidosis; antilipemic; insulin-like growth factor 2;
```


XX	IGF-II; secreted acid phosphatase; SAP; beta-glucuronidase; GUS; human;
KW	chimeric; gene; da.
OS	Homo sapiens.
OS	Leishmania mexicana.
OS	Unidentified.
OS	Chimeric.
XX	
PH	Key
FT	Location/Qualifiers
FT	1..2169
FT	/tag= a
FT	/product= "L. mexicana SAP-human beta-GUS-IGF-II chimeric protein"
FT	1..96
FT	/tag= b
FT	/note= "Leishmania mexicana SAP peptide"
FT	97..1806
FT	/tag= c
FT	/product= "Human mature beta-GUS protein"
FT	1807..1815
FT	/tag= d
FT	/note= "Linker"
FT	1816..2169
FT	/tag= e
FT	/note= "IGF-II sequence"
PN	MO200287510-A2.
PD	
PD	07-NOV-2002.
XX	
XX	30-APR-2002; 2002MO-US013835.
XX	
PR	30-APR-2001; 2001US-0287531P.
PR	10-JUL-2001; 2001US-0304609P.
PR	15-OCT-2001; 2001US-0329461P.
PR	23-JAN-2002; 2002US-0351276P.
XX	
PA	(SYMB-) SYMBIONTICS INC.
PI	
PI	Lebowitz JH, Beverley SM;
DR	WPI; 2003-111838/10.
DR	P-PSDB; AAE33322.
XX	
XX	New targeted therapeutic that is active in a mammalian lysosome binds an extracellular domain of human cation-independent mannose-6-phosphate receptor, useful for treating metabolic diseases such as lysosomal storage disease.
PT	
PT	
PT	
XX	
PS	Example 1, Fig 3; 60pp; English.
XX	
CC	The invention relates to targeted therapeutic comprising a therapeutic agent that is active in a mammalian lysosome, and a means for binding an extracellular domain of human cation-independent mannose-6-phosphate receptor in a mannose-6-phosphate independent manner. The invention is useful for treating metabolic diseases such as lysosomal storage disease, e.g. Gaucher's disease, Pompe's disease, Hurler's syndrome, Niemann-Pick's disease, mannosidosis, fucosidosis, Schindler's disease, mucopolysidosis, cystinosis, Batten's disease, prosaposin, or infantile neuronal ceroid lipofuscinosis. The invention is also useful in gene therapy. The present sequence is a DNA encoding Leishmania mexicana secreted acid phosphatase (SAP) peptide, human mature beta-GUS (glucuronidase), linker and IGF-II chimeric protein. This sequence is used in the exemplification of the invention
XX	
XX	Sequence 2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;

Query Match	10.3%	Score 174;	DB 8;	Length 2169;
Best Local Similarity	52.2%;	Pred. No. 2.8e-43;		
Matches 493;	Conservative 0;	Mismatches 430;	Indels 21;	Gaps 4
737	TGGAACCTGAAAAGACGAGTCACTCTGGAGATCGGAATTCAGACGATCAGCTGGAGACG	796		

Db	941	TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGGAATCCGCACTGTGGCTGTACCA	1000
Oy	797	AGAAAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCACGAG	856
Db	1001	AGAGCCAGTTCTCATCAATGGGAAACCTTTCTATTCTTCCAGGGTGTCAACAGCATGAGG	1060
Oy	857	AAATCCCCGTCTGTGGGCAAGGCACTTTTATCATTTGATGATPAAAAGCTTCAACCTTC	916
Db	1061	ATGCGGACATCCGAGGGAAGGGCTTGCATGCGCGCTGTGGTGAAGGACTTCAACCTGC	1120
Oy	917	TGAATGGGATCAACGCAATTTCTTTCAGGACCTTCTCACTATCTCTTACAGTGAAGTGGC	976
Db	1121	TTCCCTGGCTTGTGTCCAAAGCTTTTCGGTACAGGCACTACCCCTTATGCAAGGAAGTGA	1180
Oy	977	TGGAATCTTGCCGACAGACTGGGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTGTCA	1036
Db	1181	TGCAAGATGTGTGACCGGTATGGGATTTGTGTGATCATGATGAGTGTCTCCGGGTGGTCTGG	1240
Oy	1037	CAAGGTACCA---CTTACATCTCCGAGACTGAGAAAGATAGAGAAACAACATTAAGAA	1093
Db	1241	CGCTGCCGCACTTTTCAACAAGTTCTTCTGTATCACACATGCAAGTATGGAAGAG	1300
Oy	1094	TGATGCAAGACACAGAAGAACCATCCAGTGTGATCATGTGGAAGTGTGGCAACGAACAG	1153
Db	1301	TGTGTCGTAGGGAACAAAGAACACCCCGCGGTGCGATGTGTGTCTGTGGCAACGAGCTGT	1360
Oy	1154	AGTCCAAACCATCCAGAGCGCGGAGGCTTCTTCAAGGCCCTTATGACATCGCAATGAAA	1213
Db	1361	CGTCCACCTTAGAATCTGTGTGCTACTTGAAGATGATGTATGCTCACACCAATCTT	1420
Oy	1214	TGATTCGAACACGCGCCGTTGTCTATGTGTGACATGATGAGACGCAACGAGAGAAACAA	1273
Db	1421	TGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATGACG-----	1468
Oy	1274	GAGACGTGTGGCTGAAATCTTGACATCGTCTGTGTGAACAGGTACTACGGCTGTGCA	1333
Db	1469	CAGCAAGGGGGGCTCCGTAATGTGATGTGATCTGTGTGAACAGCTACTACTTGTGTATC	1528
Oy	1334	TCTATTCAGGGAAGGATGAAAGAAAGACTTCAAGCTCTGAAAAAAGACATGAAAGAGCTCT	1393
Db	1529	ACGACTAGGGGCACTGTGAATGTATTAAGCTGACAGCTGGCCACCCAGTTTGAAGACTGT	1588
Oy	1394	ATGCAAGGCACAGAAAGCCCATCTTTGTCAACAGATTCGGTGCGGACGCGATGCTGGCA	1453
Db	1589	ATAAGAAATGTCAAGAACCCATTAATTCAGAGCGAGTATGAGGACAGAAACGATTTGCAGGGT	1648
Oy	1454	TCCACTACGATTCACACTCAATGTCTCCGAGAGTATCCAGAGAGAGCTGTGAAAG-	1512
Db	1649	TTTACCCAGATTCACACTCTGTATTTTCACTGAAGAGTACAGAAAGTCTCTGAGCAGT	1708
Oy	1513	ACGATCAGGCTCCTTTTGAAGAAA-----AGACTACATCATCGAACAACAGTGTGGCCT	1567
Db	1709	ACCATCTGGGCTGTGATCAAAAAACGAGAAATATGTGGTTGAGAGCTCATTTGGAAAT	1768
Oy	1568	TTTGAGATTTTAAAGACTCTCAGAAATGTGAAGAACCCATTTCTCAACACAAAGGTGTT	1627
Db	1769	TTTGCCGATTTCTATGACTGAAACAGTCAACCGACGAGAGTGTGTGGGAAATAAAAAGGGGATCT	1828
Oy	1628	TCAACAAAGACAGACAAACCCAACTCGTGTCTATGTACTGAGA 1671	
Db	1829	TCACTGGCAGAGACAAACAAAAAGTGCAGCTTCTTTTGGCA 1872	

AC	ADP47502;
XX	
XX	26-FEB-2004 (first entry)
DT	
XX	
DE	Human beta-glucuronidase and IGF-II fusion protein cDNA SEQ ID NO:5.
XX	

KM underglycosylated targeted therapeutic; human; lysosome;
 KM lysosomal targeting domain;
 KM cation-independent mannose-6-phosphate receptor;
 KM underglycosylated therapeutic fusion protein; nephrotropic;
 KM enzyme replacement therapy; cell therapy; gene therapy;
 KM lysosomal storage disease; metabolic disease; enzyme defect;
 KM Pompe disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;
 KM Gaucher disease; Krabbe disease; Wolman disease; Hurler syndrome;
 KM Hunter syndrome; Sly syndrome; Schindler disease;
 KM infantile sialic acid storage disease; Batten disease;
 KM infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
 KM congenital glycosylation disorder; fusion protein; gene; ss.
 XX Synthetic.
 OS Homo sapiens.
 PN WO2003102583-A1.
 PD 11-DEC-2003.
 PD 29-MAY-2003; 2003WO-US017211.
 PF 29-MAY-2002; 2002US-0384452P.
 PR 05-JUN-2002; 2002US-0386019P.
 PR 06-SEP-2002; 2002US-0408816P.
 PR 16-OCT-2002; 2002US-00272531.
 PR 06-FEB-2003; 2003US-0445734P.
 PA (SYMB-) SYMBIONTICS INC.
 PI Lebowitz JH, Beverley SM, Sly WS,
 DR WPI: 2004-035473/03.
 DR P-PsDB; ADR47503.
 PT Novel underglycosylated targeted therapeutic comprising therapeutic agent
 PT cation-independent mannose-6-phosphate receptor, for treating Pompe
 PT disease.
 XX Example 1; SEQ ID NO 5; 137bp; English.
 XX The present invention describes an underglycosylated targeted therapeutic
 CC (GT) which comprises: (a) a therapeutic agent that is therapeutically
 CC active in human lysosome; and (b) a lysosomal targeting domain that binds
 CC an extracellular domain of human cation-independent mannose-6-phosphate
 CC receptor (CM) and (i) does not bind a mutin in which amino acid 1572 of
 CC CM is changed from isoleucine to threonine; and (ii) binds the mutin
 CC with dissociation constant at least ten times the dissociation constant
 CC for binding CM. Also described: (1) an underglycosylated therapeutic
 CC fusion protein (FP) comprising a therapeutic domain and a subcellular
 CC targeting domain that binds to an extracellular domain of a receptor on
 CC an exterior surface of a cell, and upon internalisation of the receptor,
 CC permits localisation of the therapeutic domain to a subcellular
 CC compartment where the therapeutic domain is therapeutically active; and
 CC (2) production of GT. GT has nephrotropic activity, and can be used in
 CC enzyme replacement therapy, cell therapy and gene therapy. The FP can be
 CC used for treating a lysosomal storage disease patient by administering FP
 CC to the patient. GT can be used for treating a patient by identifying a
 CC targeting moiety that binds CM in a mannose-6-phosphate independent
 CC manner, synthesizing GT comprising a therapeutic agent that is
 CC therapeutically active in a mammalian lysosome and a targeting moiety
 CC that binds CM in a mannose-6-phosphate independent manner and
 CC administering GT to the patient, where the targeting moiety is identified
 CC by screening a nucleic acid or peptide library. GT is useful for treating
 CC metabolic disease, lysosomal storage diseases and associated enzyme
 CC defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry
 CC disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler
 CC syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile
 CC sialic acid storage disease, Batten disease, infantile neuronal ceroid
 CC lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders
 CC of glycosylation. The present sequence is used in the exemplification of
 CC the present invention.

XX SQ Sequence 2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;
 Query Match 10.3%; Score 174; DB 12; Length 2169;
 Best Local Similarity 52.2%; Pred. No. 2.8e-43;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
 QY 737 TGGAACTGAAAAGACAGATGACACTGTGACATCGGAATGAGAACGATGACCTGGACG 796
 DB 941 TGGGGCTGTGTCTGACTTCTTACACACTCCCTGTGGGATCCGACTGGCTGTACCA 1000
 QY 797 AGAAGAGGCTTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAAGCAGAG 856
 DB 1001 AGAGCAGTCTCTCATCAATGAGAAACCTTTCTATTCCACGGTGTCAACAGCATGAGG 1060
 QY 857 AATTCCCGTCTTGGGGCAGGGCACCTTTATTCATTGATGATTAATAACTTCAACCTTC 916
 DB 1061 ATGGGAGCATCCGAGGGAGGGCTTGCAGCTGGCGCTGTGTGAMGACCTTCAACCTGC 1120
 QY 917 TGAAGTGATCAAGCGCAATCTTTCAAGACCTCTCACTATCTTACAGTGAAGTGGC 976
 DB 1121 TTGCTGTGCTGTGTGTCACAGCTTTCCGTACAGCACTTACCCTTATGCAAGAGAACTGA 1180
 QY 977 TGGATCTTGGCCAGACAGACTCGAAATCTTGTGATAGACGAAGCCCGCAGCTTGATCA 1036
 DB 1181 TGCAAGATGTGACCGCTATGAGATTGTGTGATCATGATGATGTCCCGGCTGGTCTGG 1240
 QY 1037 CAAAGTACCA--CTACAATCCCGAGACTCGAAGATGACGAAAGACAAACATTAAGAA 1093
 DB 1241 CGTGCAGAGTTCTTCAACAAAGTTTCTGTGCATCACACATGCAAGGTGATGAAAG 1300
 QY 1094 TGATCGACAGACACAGAAACCATCCAGTGTGATCATGTGAGTGTGCGAAGCAACGAG 1153
 DB 1301 TGTGTGCGTGAAGGACAGAAACCAACCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1360
 QY 1154 AGTCCAACTATCAGACGCGGAGGGTTCCTTCAAGCCCTTTATGAGACTGCCAATGAA 1213
 DB 1361 CGTCCCACTGAAATCTGTGCTGCTACTTGAAGATGATGATGCTCACACCAATCTCT 1420
 QY 1214 TGGATCGAACACGCCCGCTGTGTGATGATGATGATGATGATGATGATGATGATGATG 1273
 DB 1421 TGGACCTCTCCCGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1468
 QY 1274 GAGACGTGGCCCTGAAGACTTTCGACATCGTGTGTGAACAGGATCTACGCTGTGACA 1333
 DB 1469 CAGACAGAGGGGCTCCGATGTGATGTGATCTGTTTGAACAGCTACTTCTTGTGTATC 1528
 QY 1334 TCTATCAGGAAGGATAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1393
 DB 1529 ACGACTACGGGACCTGGAATGATTCAGCTGACGCTGGCACCACCAAGTTGAGAACTG 1588
 QY 1394 ATGCAAGCACAAGAAAGCCCATCTTTGTACAGAAATTCGTGTGCGGACCCGATCTGTG 1453
 DB 1589 ATAAAGAAATATCAGAAAGCCCATTTATTCAGACCGGATATGAGCAGAAACGATTC 1648
 QY 1454 TCCACTACGATCCCACTCAATGTTCTTCGAAGATGACCAAGCAGAGCTCGTTGAAAAG- 1512
 DB 1649 TTCAACAGGATCCACTCTGTGATGTTCACTGAAGATCACAAGAAAGTCTGTGAGCACT 1708
 QY 1513 ACGATCAGGCTCTTTTGAAGAA--AGACTACATCATCGGAACACAGCTGTGGGCT 1567
 DB 1709 ACGATCTGGGCTGTGATCAAAAACGACAGAAATATGTGGTTGAGAGCTCATTTTGA 1768
 QY 1568 TTGCAGATTTTAAGACTCTCAGAAATGTGAGAAAGCAATCTCTCAACCAAGGCTGTT 1627
 DB 1769 TTGCGGATTTATGATGATGAAAGTCACCGAGAGTGTGGGAAATTAAGAGGATCT 1828
 QY 1628 TCACAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
 DB 1829 TCACCTGGCAGAGACACCAAAAGTGAAGCTTCTTTTGA 1872

RESULT 8

ADJ58608	
ID	ADJ58608 standard; DNA; 2169 BP.
XX	
AC	ADJ58608;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Fusion DNA used in the exemplification of the invention.
XX	
KW	Targetted therapeutic; lysosome storage disease; gene therapy; human; IGF-II; insulin-like growth factor 2; beta-glucuronidase; GUS; chimeric; gene; ds.
XX	
OS	Homo sapiens.
OS	Chimeric.
OS	Unidentified.
XX	
PH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2169
FT	/*tag= b
FT	/product= "Fusion protein"
FT	1..69
FT	misc_feature
FT	/*tag= a
FT	/note= "Signal peptide DNA"
FT	70..1956
FT	/*tag= c
FT	/note= "Human mature P-glucoronidase DNA"
FT	1957..1965
FT	misc_feature
FT	/*tag= d
FT	/note= "Encodes an amino acid bridge"
FT	1966..2166
FT	/*tag= e
FT	/note= "Human IGF-II DNA"
XX	
PN	US2004006008-A1.
XX	
PD	
ED	08-JAN-2004.
XX	
PF	16-OCT-2002; 2002US-00272483.
XX	
PR	30-APR-2001; 2001US-0287531P.
PR	10-JUL-2001; 2001US-0304609P.
PR	15-OCT-2001; 2001US-0329461P.
PR	23-JAN-2002; 2002US-0351276P.
PR	30-APR-2002; 2002US-00136841.
PR	29-MAY-2002; 2002US-0384452P.
PR	05-JUN-2002; 2002US-0386019P.
PR	06-SEP-2002; 2002US-0408816P.
XX	
PA	(SYMB-) SYMBIONICS INC.
XX	
PI	Lebowitz JH, Beverly SM;
XX	
DR	WPI; 2004-081736/08.
DR	P-PSDB; ADJ58609.
XX	
PT	New targeted therapeutic comprising a therapeutic agent that is active in
PT	mammalian lysosome, and means for binding a domain of human cation-
PT	independent mannose-6-phosphate receptor, useful for treating lysosome
PT	storage disease.
XX	
PS	Example 1; SEQ ID NO 5; 46p; English.
XX	
CC	The invention relates to a targetted therapeutic which comprises a
CC	therapeutic agent that is active in a mammalian lysosome and means for
CC	binding an extracellular domain of human cation-independent mannose-6-
CC	phosphate receptor in a mannose-6-phosphate-independent manner. The
CC	targetted therapeutic, therapeutic fusion protein and methods of the
CC	invention are useful for treating lysosome storage diseases. The
CC	invention is also useful in gene therapy. The present sequence is a
CC	fusion DNA used in the exemplification of the invention. This DNA encodes
CC	a fusion protein which comprises a signal peptide sequence, a human
CC	mature beta-glucuronidase (GUS) sequence, a bridge of three amino acids
CC	

CC	and a human IGF-II sequence.
XX	
Sequence	2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;
Query Match	10.3%; Score 174; DB 12; Length 2169;
Best Local Similarity	52.2%; Pred. No. 2.8e-43;
Matches 493; Conservative	0; Mismatches 430; Indels 21; Gaps 4

OY	737	TGGAACTTTGAAAAAGACAGATGCACTCTGGACATCTGGAAATCGAAGACGATCAAGCTGAGGACG	796
Db	941	TGGGGCTGTGTCTGACTTCTTACACACTCTCTGTGGGAGATCGACCTGTGGCTGTCA	1000
OY	797	AGAAAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGGAAAGGGCTTTGGAAAGCAGAG	856
Db	1001	AGAGCGAGTCTCATCAATGGGAAACCTTTCTATTTCCAGGTGTCAACAAGCATGAG	1066
OY	857	AATTCCTCGTTCTGGGCGAGGGCACCTTTATTCATTGATGATTAAGAACTTCAACTTC	916
Db	1061	ATGGCGACATCCGAGGGAGGGCTTGCATCGGCGCTGTGGTGAAGGACTTCAACTGCG	11220
OY	917	TGAAGTGATCAAGCGCAATCTTTCAAGGACCTCTCACTATCTTTACAGTGAAGTGTC	976
Db	1121	TTCCGTGCTTGTGTGCACACGCTTTCCGTACAGCACTACCCCTATGCAAGAGAAATGA	1180
OY	977	TGATCTTTGCCGACAGACTTCGGAATCCCTTGATATGACGAAGACCCCGACGCTTGATCA	1033
Db	1181	TGCAGATGTGTGACCGCTATGGAGTTGTGTGATCGATGAGATGTCCGGCTGGGTCTGG	1244
OY	1037	CAAGGTACCA--CTACATCCCGAGACTCGAAGATGACGAAGAACAACTTAAGAGAA	1093
Db	1241	CGTGCAGCATGTTCTTCAACAACGTTTCTGTCAACACATGACAGGTGATGGAAGAG	1300
OY	1094	TGATCGACAGACACAAGAACCATCCCACTGTGATCATATGTGGAGTGTGGCCAAACCA	1155
Db	1301	TGGTGCGTAGGGAACAAAGAACACCCCGGGTCGTGATGTGTGTGTGCCAAACGAGCTTG	1366
OY	1154	AGTCAACCATTCACAGACGCGGAGGGTTCCTCAAGCCCTTATGAGACTCGCAATGAAA	1215
Db	1361	CGTCCACCTTGAATCTGTGTGGCTACTACTTGAAGATGTGATGTCTCAACCAATCTT	1422
OY	1214	TGATTCGAACACGCCCCGTGTGATGTGATGACATGATGGAACGCAACAGACAGAGAACAA	1273
Db	1421	TGACACCCCTCCGGGCTGTGACCTTTTGACACCACTTAATGACAG-----	1466
OY	1274	GAGACGTGTGCGCTGAAGTACTTTCACATCGTCTGTGTGAACAAGTACTCAAGCGTGTGA	1333
Db	1469	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTCTTGGTATC	1528
OY	1334	TCTATCAGGGAGGATGGAAGAAGACTTCAAGCTCTGGAAAAAGACATAGAAGCTCT	1392
Db	1529	ACGACTAGGGACCTGTGATGTGATTCAGCTGCAGCTGGCCACCCAGTTTGAAGACTGT	1588
OY	1394	ATGCAAGGCAAGAAGGCCATCTTTTTCACAGAAATTCGTGTGCGGACGCAATAGCTGCA	1453
Db	1589	ATTAGAAGTATCAGAAAGCCATTATTCAGACGAGTATGAGGACGAACATTTGAGAGGT	1644
OY	1454	TTCACTACGATTCACCTCAAAATGTTCTTCGAAGAGTACCAAGACAGAGCTCGTTGAAG-	1512
Db	1649	TTCAACAGGAATCCACCTGTGATGTTCCTACTGAAGAGTACCAAAAAAGTCTGTAGAGCAGT	1708
OY	1513	ACGATCAAGGCTCTTTTGAAGAA-----AGACTATCATATGGGAACACAGCTGTGGGCT	1567
Db	1709	ACCATCTGGGCTGTGATTCAAAAACGCAAAAAATATGTGGTTGGAGGCTCATTTTGAAT	1766
OY	1568	TTGACAGTTTTTAAGACTCTCAGATGTGGAAGAACCACTTCFAACACAGAGGTGTTT	1627
Db	1769	TTTGCAGATTTATATCTGAACAGTCAACCGAGAGAGTGTCTGGGAATTAAGAGGAGTCT	1828
OY	1628	TCACAAAGACAGAACCAACCAACTCTGTCTCATGTACTGAGA	1671
Db	1829	TCACCTGGCAGAGCAACCAAAAAAGTGAGGCTTCTTTTSCGA	1872

RESULT 9
AAB06386
ID AAB06386 standard; DNA; 1956 BP.
XX
AC AAB06386;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human beta-glucuronidase (GUS) DNA.
XX
KM Human; adeno-associated viral expression vector; AAV; gene therapy;
KM lysosomal storage disease; LSD; mucopolysaccharidosis VII; MPS VII;
KM Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1956
FT /tag= a
FT /product= "Human beta-glucuronidase (GUS) "
XX
PN MO200136603-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US031688.
XX
PR 17-NOV-1999; 99US-0166097P.
XX 30-JUN-2000; 2000US-0215430P.
XX
PA (AVIG-) AVIGEN INC.
XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI Podsakoff G, Watson G, Couto LB, Yang B;
XX
DR WPI; 2001-343814/36.
XX P-PSDB; AAE02443.
XX
PT Use of recombinant adeno-associated virus, comprising gene encoding a
PT protein defective or missing in lysosomal storage disease, in the
PT manufacture of a medicament for treating the lysosomal storage disease.
XX
PS Disclosure; Page 77-80; 97pp; English.
XX
CC The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful in
CC the manufacture of a medicament for treating lysosomal storage disease
CC e.g., mucopolysaccharidosis VII (MPS VII). MPS VII (Sly Syndrome) is due
CC to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) which
CC aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in gene
CC therapy. The present sequence is human GUS DNA. This sequence is used in
CC AAV constructs
XX
SQ Sequence 1956 BP; 437 A; 533 C; 563 G; 423 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 4; Length 1956;
Beet Local Similarity 52.1%; Pred. No. 8.4e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAATAAGACGATCACTTGTGACATCGAATCGAAGCATCAGCTGGGACG 796
DB TGGGGCGTGTGTCTGACTTCTACACACTCCCTGTGGGATCGGACGTGGCTGCACCA 997
QY 797 AGAAGGCTCTATCTGAACGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCAGAG 856
DB 998 AGAGCAGTCTCTCATCAATGGGAAACCTTCTATTTCCACGGTGTCAACAGCATGAGG 1057
QY 857 AATTCCCGCTTGGGGGACGACCTTTATCCATGATGATTAATAAGCTTCAACCTTC 916
DB 1058 ATGCGGACATCTCGAGGAGGCTTGCATGGCGCTGTGTGAAGGACTTCAACCTGC 1117
QY 917 TGAAGTGATCAACGGAATCTTTCAGGACCTTCATCTACTTCAAGTGAAGAGTGGC 976

DB 1118 TTGCGTGGCTTGTGCGCAAGACCTTTCCGTACACAGCACTACCCCATGACAGAGAAATGA 1177
QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGACCTGGTATCA 1036
DB 1178 TGCAGATGTGTGACCGCTATGGGATTTGTGTATCATGATGATGTCTCCGGCGGCTGG 1237
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTGAGAGATGACGAAGAACATTAAGAGAA 1093
DB 1238 CGCTGCCGAGTCTTTCACCAACGTTTCTCTGCATCACCATGATGAGGTATGGAAGAG 1297
QY 1094 TGAATCAGACAGACAGAAGAACATCCAGTGTATCATGTGAGTGTGGCAGAACAG 1153
DB 1298 TGTGTGCGTGAAGAGAACAAACACCCCGGCTGTGATGTGTGTGGCAGAGCCCTG 1357
QY 1154 AGTCCACATCCAGACGGCGAGGCTTTCTTCAAAAGCCCTTTATGAGCTGCCAATGAAA 1213
DB 1358 GGTCCACCTTGAATCTGTGCTGTACTTATGGAAGATGTATGCTCACACCAATGCT 1417
QY 1214 TGAATCGAACAAGCGCCGTTGTGATGTGAGACATGATGAGCACACAGACGAGAGAA 1273
DB 1418 TGGACCCCTCCCGGCTGTGTACTTTGTGACCACTTAATGATGACG----- 1465
QY 1274 GAGACGTGGCGCTGAAAGTACTTGCACATCGTCTGTGTAACAGGTACTACGCTGTGTA 1333
DB 1466 CAGACAGGGGGGCTCCGTATGTGATGTGATCTGTGTAACAGGTACTACCTTGTGTATC 1525
QY 1334 TCTATCAGGAAAGATGAAAGAGACTTCAAGCTCTGTGAAAAGACATAGAAGCTCT 1393
DB 1526 ACGACTACGGGACCTGAGTGTGATTCAGCTGCAGCTGGCCACCCAGTTTGAAGACTGT 1585
QY 1394 ATGCAAGGACAGAAAGCCATCTTTGTGCACAGAAATTCGTGCGGACCGATAGCTGCA 1453
DB 1586 ATTAAGAGTATCAGAAAGCCATTATTCAGACGAGATAGAGCAGAAACGATGAGGCT 1645
QY 1454 TCCACTACGATCAACCTTCAATGTTCTCCGAAAGATCAACAGACAGAGCTGTTGAAA 1512
DB 1646 TTCAACAGATTCACCTGTATGTTCACTGAAGAGTACCAAAAAGTCTGTAGAGCAGT 1705
QY 1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATATATGGGAACACACGTGTGGCT 1567
DB 1706 ACCATCTGGGTCTGAGTCAAAAACGACAGAAATATGTGTGAGAGCTCATTTGGAATT 1765
QY 1568 TTGCAGATTTTAAGCTCCTCAGATGTGAGAGACCATTTCTAACCAAGAGGTGTTT 1627
DB 1766 TTGCGGATTTATGATCTAATCAAGTCAACGAGAGAGTCTGGGAATTAAGAGGATCT 1825
QY 1628 TCACAAGAGACAGAACCAACCACTGTTGCTCATGTACTGAGA 1671
DB 1826 TCACCTGGCAGAGACAAACAAAAGTGACGCTTCTTTTTCGA 1869

RESULT 10
ABK83932
ID ABK83932 standard; cDNA; 2191 BP.
XX
AC ABK83932;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #503.
XX
KM Human; ss; granulocytic cell; DNA chip; bacterial infection;
KM viral infection; parasitic infection; protozoal infection;
KM fungal infection; sterile inflammatory disease; poxvirus;
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
KM adult respiratory distress syndrome; inflammatory bowel disease;
KM Crohn's disease; ulcerative colitis; periodontal disease;
KM granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX

[illegible]

QY	917	CGAATGATCAACCGCAATTCCTTTCAGACCTCTCACTATCTTACAGTGAAGTGGC	976
Db	1144	TTCCGTGGCTTGGTGCACACGCTTTCGTACCGACCTTACCTTATGCAGAGAACTGA	1201
QY	977	TGATCTTTCGCGACAGACTCGGAATCTTGTGTATGACGAAGCCCGCACGTTGTATCA	1036
Db	1204	TGCAGATGTGTGACCGGCTATGGGAATGTGTGTATGATGATGATGTCTCCGGCGTGGC	1263
QY	1037	CAAGGTACCA---CTACATATCCCGAGACTCAGAGATATGCAGAGACAACTAAGAAGA	1097
Db	1264	CGTCCCGCAATTCCTTCAACAACGTTTCTCTGCATCAACAATGACAGGTGTATGAAGAAG	1322
QY	1094	TGATGCACAGACACAAGAACCATCCCACTGTGATCATGTGAGTGGAGTGCAGAACGACAG	1155
Db	1334	TGTGTGCTGAGGACAGAAACACCCCGGCTGTGTATGTGTGTGTGTGTGTGTGTGTGTGT	1388
QY	1154	AGTCCAAACCATTCAGACCGCGAGGGTTCCTTCAAGCCCTTATATGAGACTGCCAATGAA	1213
Db	1384	CGTCCCACTTGAATCTGT	1444
QY	1214	TGATTCGAACACGCCCCGT	1273
Db	1444	TGGAACCCCTCCCGGCTGT	1491
QY	1274	GAGACGTGGCGCTGAAGTCTTCGACATCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1333
Db	1482	CAGACAGAGGGGCTTCGT	1551
QY	1334	TCTATCAGAGGAGATGAGAGAGACTTCAAGCTCTGGAAGAAAGACATAGAGAGCTCT	1393
Db	1552	ACGACTACGGGCACTGTGATGT	1611
QY	1394	ATGCAAGGCAAGAAAGCCCATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1451
Db	1612	ATAGAAATTAACAGAAAGCCATTTATTCAGACCGAGTATGAGACAGAAACGATTCGAGGGT	1671
QY	1454	TCCACTACGATCCACCTCAAAATGTTCTCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG	1512
Db	1672	TTTACACAGATCCACCTTGT	1731
QY	1568	TTGCAAGATTTTAAAGCTCTCTCAGATGTGAGAAAGCCATTTCTAACAAGGGTGT	1622
Db	1792	TTGCGGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1851
QY	1628	TCACAAGAGACAGAACCCCAAACTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1671
Db	1852	TCACTTCGCGAGACACACCAAAAGTGTGAGGCTTCTTTTGTGGA	1895
RESULT 11			
ADD29602			
ID ADD29602 standard; mRNA: 2191 BP.			
XX ADD29602;			
DT 15-JAN-2004 (first entry)			
XX Human tumour suppressor mRNA SEQ ID NO:55.			
DE ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.			
XX KW Homo sapiens.			
XX OS			
XX PN MO2003058201-A2.			
XX PD 17-JUL-2003.			
XX 31-DEC-2002; 2002WO-US041825.			
PF			

XX 31-DEC-2001; 2001US-0345317P.
 PR (QUAR-) QUARK BIOTECH INC.
 PA (CLEV-) CLEVELAND CLINIC FOUND.
 PI Feinstein E, Gudkov AV;
 DR MPI; 2003-598393/56.
 XX
 PT Diagnosing cancer comprises determining the polypeptide or polynucleotide
 PT levels e.g., hepatic lipase, in a sample from a subject, where a higher
 PT level compared to that in a subject free of cancer is indicative of
 PT cancer.
 XX
 PS Disclosure; SEQ ID NO 55; 272pp; English.
 XX
 CC The invention relates to a novel method for diagnosing a cancer in a
 CC subject, the method comprises determining, in a sample from the subject,
 CC the level of at least one polypeptide, where a higher level of the
 CC polypeptide compared to the level of the polypeptide in a subject free of
 CC cancer is indicative of cancer. The polypeptide is selected from any of
 CC the polypeptides encoded by the polynucleotides listed in the
 CC specification and polypeptides which are at least 70% homologous to the
 CC polypeptides. The method of the invention has cytostatic activity, and
 CC may have a use in gene therapy. The method is useful in identifying
 CC markers specific for one or several types of cancer, depending on the
 CC tissue origin, which may be used in numerous diagnostic and prognostic
 CC applications as well as cancer type-specific targets for therapeutic
 CC intervention. The compounds that modulate the activity of a tumor
 CC suppressor gene are useful in the treatment of cancer or as anti-cancer
 CC drugs. The present sequence represents a polynucleotide of the invention.
 XX
 SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 10; Length 2191;
 Best Local Similarity 52.1%; Pred. No. 9e-43;
 Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGGACATCGGAATCGAATCGATAGCTGGAGC 796
 DB 964 TGGGGCTGTGTCTGATCTTCAACATCTCTGGGGATCCGACTGTGCTGTACCA 1023
 QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGGCTTGGAAACAGAG 856
 DB 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTCCACGGTGTCAACAAAGCATGAG 1083
 QY 857 AATTCCTCCGTCTGGGGCAGGGCACTTTTATCATGTGATGATTAAGAACTTCACTTC 916
 DB 1084 ATGCGGACATCCGAGGAAAGGGCTTCGATGCGCTGCTGTGAAGACTTCAACTGCG 1143
 QY 917 TGAAGTGAATCAACCGGAATTTCTTCAGGACCTTCACATATCTTATCAAGTGAAGTGGC 976
 DB 1144 TTGCGTGGCTGTGTGCAACGCTTTTCCGACGACCATCACTTATGAGGAGTGA 1203
 QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTATAGACGAACCCCGACGTTGGTATCA 1036
 DB 1204 TGCAGATGTGTACCGCTATGGGATTTGTGTATCATGATAGTGTCCCGCTGGGCTCG 1263
 QY 1037 CAAGGTACCA---CTAACAATCCCGAGACTCAGAAGATAGCAGAAGCAATAGAAGA 1093
 DB 1264 CCTGCGCGGAGTTCTTCAACAACGTTTCTGCAATCACCACATGCGAGTATGAAGAAG 1323
 QY 1094 TGAATGACAGACAAAGAAACCATCCAGTGTATCATGTGAGTGTGGGAAACGAACGAG 1153
 DB 1324 TGTGTGTGAGGAGCAAGAAACACCCCGCGTGTATGTGTGTGTGGCAACGAGCGCTG 1383
 QY 1154 AGTCCAAACCATCCAGACGCGAGAGGTTTCTTCAAAAGCCCTTATATAGATGCGCAATGAA 1213
 DB 1384 CGTCCACCTTAATATGTGCTGTGCTACTTGAATGTGTATCGCTCACACCAATCTT 1443
 QY 1214 TGGATGAAACACGCCCCGTTGTATGTGTGAGATGATGAGCCACACGAGAGAAACAA 1273

DB 1444 TGAACCCCTCCCGGCTGTGACCTTTGTGACCACTTAATATGACAG----- 1491
 QY 1274 GAGACGTGGCGCTGAGTACTTGCACATCGTCTGTGTAACAGGATACGGCTGTACA 1333
 DB 1492 CAGACAAAGGGGCTCCGCTATGTGATGTGATGTGTAACAGCTACTACTTGTGTATC 1551
 QY 1334 TCTATCAGGAGAAAGATAGAAAGAACTTCAAGCTTGGAAAAAGACATAGAGAGCTCT 1393
 DB 1552 ACGACTACGGGCACTGAGATGTGATTCAGCTGCGAGCTGCCACCAAGTTTGAAGACTGT 1611
 QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGTGCGGACCGCATAGCTGCA 1453
 DB 1612 ATTAAGAAATATCAGAAAGCCCATTTATTCAGACCGGATATGAGCAGAAACGATTTGCGAGGT 1671
 QY 1454 TCCACTACGATCCACTCAATGTCTTCCGAAGATCCAGCAGAGCTCGTTGAAAG- 1512
 DB 1672 TTCACGAGATCCACTCTGATGTTCAGTGAAGATCAACAAAGTCTGCTAGAGCAT 1731
 QY 1513 ACGATCAGGCTCTCTTTGAAAAA-----AGACTATCATCGAACAACAGCTGTGGCCT 1567
 DB 1732 ACGATCTGGGTCTGGATCAAAAACGAGAAATATGTGTGAGAGCTCATTTGGATTT 1791
 QY 1568 TTGCAATTTTAAAGCTCTCTCAGAAATGTGAGAAAGCCATTTCTAACCAACAGGCTTTT 1627
 DB 1792 TTGCGATTTTATGATGATGAAACATGATCAGCAGAGAGTGTGGGAAATTAAGGGGATCT 1851
 QY 1628 TCACAAGAGACAGAACCAACCCAACTCGTGTCTCATGTACTGAGA 1671
 DB 1852 TCATCTGGCAGAGAACCAACAAAAGTGCAGGCTTCTTTTGGCA 1895

RESULT 12
 ADG89395
 ID ADG89395 standard; DNA; 2191 BP.
 XX
 AC ADG89395;
 XX
 DT 11-MAR-2004 (first entry)
 XX

DE Cancer detection method related gene #58.
 KW ds; cancer; gene expression;
 KW estrogen receptor-positive invasive breast cancer.
 OS Homo sapiens.
 FN WO2003078662-A1.
 XX
 PD 25-SEP-2003.
 XX
 PF 12-MAR-2003; 2003WO-US007713.
 XX
 PR 13-MAR-2002; 2002US-0364890P.
 PR 18-SEP-2002; 2002US-0412049P.
 XX
 PA (GENO-) GENOMIC HEALTH INC.
 PI Baker JB, Cronin MT, Kiefer MC, Shak S, Walker MG;
 XX
 DR MPI; 2003-767536/72.
 XX
 PT Predicting clinical outcome for a patient diagnosed with cancer comprises
 PT determining the expression level of one or more genes, and compared to
 PT the amount found in a reference cancer tissue set.
 XX
 PS Disclosure; SEQ ID NO 343; 198pp; English.
 XX
 CC The invention relates to a method of predicting clinical outcome for a
 CC patient diagnosed with cancer by determining the expression level of one
 CC or more genes, or their expression products, selected from p53BP2,
 CC cathepsin B, cathepsin L, Ki67/MIB1, and thymidine kinase in a cancer
 CC tissue obtained from the patient, normalized against control gene(s), and
 CC compared to the amount found in a reference cancer tissue set. The

QY 737 TGGAACTTGGAAAAAGACGATACACTCTGAGATCGGAATCAGAAAGATCAGCTGGAGCG 796
 DB 964 TGGGGGCTGTGTCTGCTTCTACACATCCCTGTGGGAGATCGGACCTGTGGTTCACCA 1023
 QY 797 AGAAGAGCTTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAANGCAGAG 856
 DB 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAGCATGAGG 1083
 QY 857 AATTCCCGCTTGGGGGAGGGGACCTTTTATCCATTGATGATTAAGACCTTCAACCTTC 916
 DB 1084 ATGCGGACATCCGAGGAAAGGGCTTGCATGCGCTGCTGTGTAAGACCTTCAACCTGC 1143
 QY 917 TGAAGTGATCAACGGGAATTTCTTCAGGACCTCTCATCTTCACTGTAAGAGTGGC 976
 DB 1144 TTCGCTGGCTTGTGTGCAACGGCTTTCGTACCAAGCACTAACCTTATGAGAGGAAGTGA 1203
 QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTGATGAGAGAGCCCGGACGTTGGTATCA 1036
 DB 1204 TGCAGATGTGTACCGCTATGGGATGTGTGTCATGATGATGATGATCCGGCGCTGGGCTGG 1263
 QY 1037 CAAGGTACCA---CTACAAATCCCGAGACTCAGAAAGTACAGAGCAACATTAAGAGTA 1093
 DB 1264 CCGTCCCGGAGTTCTTCAACAAAGCTTCTGTGATCACCAATGCAAGTGTGATGAGAGTA 1323
 QY 1094 TGATGACAGACAGACAAAGCAATCCAGTGTATCATGTGAGTGTGGCGAAGCAACAG 1153
 DB 1324 TGTGTGTGTGGGACAAAGACCAACCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1383
 QY 1154 AGTCCAACTATCAGACGCGGAGGGTCTTCAAAAGCCCTTATGAGTGTGTGTGTGTGTGT 1213
 DB 1384 CGTCCCACTAGATCTGT 1443
 QY 1214 TGGATGCAACAGCCCGCTTGTGATGAGATGATGAGACCAAGCAAGCAAGAGAGAAACA 1273
 DB 1444 TGGACCCCTCCGGGCTGTGACCTTTGTGAGCACTTACTACTATGAG--- 1491
 QY 1274 GAGAGTGGGCTGTAAGTACTTCAATGCTGTGTGTGAAACAGTACTACGGCTGTGTACA 1333
 DB 1492 CAGACAAAGGGGCTCGTATGT 1551
 QY 1334 TCTATCAGGGAAGATGAAAGAAAGACTTCAAGCTCTGGAAGAAACATTAAGAGCTCT 1393
 DB 1551 AGGACTACGGGACCTGGAGTGTATGACTGCAAGCTGCGCAACCAAGTTGAGAACTGGT 1611
 QY 1394 ATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGGACGGATGAGTGGCA 1453
 DB 1612 ATTAAGAGTATCAGAAAGCCCATTTATTCAGAGGAGTATGAGAGCAAGAAAGATTGCA 1671
 QY 1454 TGCACCTACGATCCACTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAAG- 1512
 DB 1672 TTCACGAGATCCACTCTGATGTTCACTGAAAGAGTACCAAGAAAGTGTGTGAGAGAGT 1731
 QY 1513 AGGATCAGGCTCTTTTGAAGAA---AGACTACATCATCGGAACACAGCTGTGGGCT 1567
 DB 1732 ACCATCTGGGTGTGATCAAAAACGCAAAAATATGTGTGTAAGAGCTCATTTGGAAT 1791
 QY 1568 TGGAGATTTTAAAGCTCTCAGAAATGTGAAAGAACCAATTCACACCAAGAGAGTGT 1627
 DB 1792 TTGCGGATTTCACTGATGAACAGTACCGAGAGAGTGTGGGAGATTAAGAGGGAGTCT 1851
 QY 1628 TCACAGAGACAGACAAACCAACTGTTGTCTCATGTACTGAGA 1671
 DB 1852 TCACCTGGCAGAGACAAACCAAAAGTGCAGCTTCTTTTGGCA 1895

RESULT 14
 ADJ62864
 ID ADJ62864 standard; DNA; 2191 BP.
 XX
 AC ADJ62864;
 XX
 DT 06-MAY-2004 (first entry)
 XX

DE Human beta-glucuronidase gene related to cancer treatment.
 XX beta-catenin; bipartite T-cell factor; Tcf; promoter construct; LRF-1;
 KW cytosolic; beta-catenin activity inhibitor; gene therapy; colon cancer;
 KW metastasis; liver; thymidine kinase; prodrug; chemotherapy;
 KW radiation therapy; surgery; human; beta-glucuronidase; gene; ds.
 OS Homo sapiens.
 PN US2003228285-A1.
 PD 11-DEC-2003.
 XX
 PF 05-MAY-2003; 2003US-00429802.
 XX
 PR 03-MAY-2002; 2002US-0377672P.
 XX
 PA (HUNG/) HUNG M.
 PA (KMON/) KWONG K Y.
 PA (ZOU/) ZOU Y.
 PI Hung M, Kwong KY, Zou Y;
 DR WPI: 2004-042209/04.
 XX
 PT Novel viral vector comprising beta-catenin/bipartite T-cell factor-
 PT responsive promoter having first and second promoter region linked to
 PT target nucleic acid sequence, useful for treat- ing colon cancer.
 PS Disclosure; SEQ ID NO 16; 114pp; English.
 XX
 CC This invention relates to a novel viral vector comprising a beta-
 CC catenin/bipartite T-cell factor (Tcf)-responsive promoter construct which
 CC contains a first promoter region having a copy of Tcf/LRF-1 binding site
 CC operatively linked to a second promoter region, and a nucleic acid
 CC sequence, where the first and second promoter regions are operatively
 CC linked to target nucleic acid sequence. The invention may be useful for
 CC the development of compounds with a cytosolic activity, through action
 CC as beta-catenin activity inhibitors, or for gene therapy. The invention
 CC may be useful for treating an individual with colon cancer which is
 CC metastasized to the liver. The treatment involves administering the
 CC vector of the invention where the nucleic acid sequence encodes a
 CC therapeutic polypeptide or thymidine kinase, a prodrug and chemotherapy,
 CC radiation, surgery or gene therapy to the individual. The present
 CC sequence is that of a therapeutic gene which may be used in the vector of
 CC the invention for the treatment of an individual with cancer.
 SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
 Query Match 10.2%; Score 172.4; DB 12; Length 2191;
 Best Local Similarity 52.1%; Pred. No. 9e-43;
 Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGGAAAAAGACGATACACTCTGAGATCGGAATCAGAAAGATCAGCTGGAGCG 796
 DB 964 TGGGGGCTGTGTCTGCTTCTACACATCCCTGTGGGAGATCGGACCTGTGGTTCACCA 1023
 QY 797 AGAAGAGCTTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAANGCAGAG 856
 DB 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAGCATGAGG 1083
 QY 857 AATTCCCGCTTGGGGGAGGGGACCTTTTATCCATTGATGATTAAGACCTTCAACCTTC 916
 DB 1084 ATGCGGACATCCGAGGAAAGGGCTTGCATGCGCTGCTGTGTAAGACCTTCAACCTGC 1143
 QY 917 TGAAGTGATCAACGGGAATTTCTTCAGGACCTCTCATCTTCACTGTAAGAGTGGC 976
 DB 1144 TTCGCTGGCTTGTGTGCAACGGCTTTCGTACCAAGCACTAACCTTATGAGAGGAAGTGA 1203
 QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTGATGAGAGAGCCCGGACGTTGGTATCA 1036
 DB 1204 TGCAGATGTGTACCGCTATGGGATGTGTGTCATGATGATGATGATCCGGCGCTGGGCTGG 1263

QY 1037 CAAGTACCA---CTACATCCCGAGACTCAGAAATAGACAGAAACAATAGAAAGA 1093
 DB 1264 CCTCCCGAGATTCTTCAACAGGTTCTCTGCATACCAATGCAGGTGATGAAAGAG 1323
 QY 1094 TGATGACAGACAGAAACCAATCCCAATGTGATCATGTGAGTGTGCGAAACGACAG 1153
 DB 1324 TGGTGTGAGGACAGAAACCAACCCCGGTGTGATGTGTGTGCGCAACGAGCTG 1383
 QY 1154 AGTCCACCATCCAGACGGGAGGTTTCTTCAAGCCCTTATATGAGTCTCCATGAAA 1213
 DB 1384 CTTCCACCATAGATCTGTGTGCTACTTCTTGAATGTGTATGTGTCTCACCAATCTCT 1443
 QY 1214 TGGATGAAACAGCCCGCTTGTATGATGATGATGATGATGATGATGATGATGATGAT 1273
 DB 1444 TGGACCTCTCCGCTGTGATCTTGTGATGATCTTGTGATGATCTTGTGATGATCTT 1491
 QY 1274 GAGACGTGCGCTGAAATCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 1333
 DB 1492 CAGACAGAGGGGCTCGTATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 1551
 QY 1334 TCTATCAGGAAAGATGAAAGACCTTCAAGCTTGGAAAAACATAGAAAGCTCT 1393
 DB 1552 AGGACTACGGGACCTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 1611
 QY 1394 ATGCAAGGACAGAAAGCCATCTTGTGACAGATTTGGTGGAGCGGATAGCTGGCA 1453
 DB 1612 ATTAGAGATGATCAGAAAGCCATTTATCAGAGGAGATGATGAGAGAAAGATTTGAGGAT 1671
 QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAAGATGACCAAGAGCTGTGAAAG- 1512
 DB 1672 TTCACGAGATCCACTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 1731
 QY 1513 AGCATGAGCTCTTTTGAAGAA---AGACTACATCATTCGAAACACGCTGTGGCCT 1567
 DB 1732 ACCATCTGGGTCTGTGATCAAAAACGCAAAAATGTGTGTGAGAGCTCATTTGGAAAT 1791
 QY 1568 TTGAGATTTTAACTCTCAGATGTGAAAGACCATTTCTCAACCAAGAGGATTT 1627
 DB 1792 TTGCGATTTCTAGTGAACGTACCGACGAGAGTGTGTGGAAATTAAGAGGATCT 1851
 QY 1628 TCACAGAGACAGAAACCAACTGTGTCTCATGTACTGTAGA 1671
 DB 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1895
 RESULT 15
 ADP10326
 ID ADP10326 standard; DNA; 2191 BP.
 XX
 AC ADP10326;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Reference mRNA sequences for marker probe #3.
 XX
 KM transplant rejection; immune system; rheumatoid arthritis; lupus;
 KM inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2004042346-A2.
 XX
 PD 21-MAY-2004.
 XX
 PF 24-APR-2003; 2003WO-US012946.
 XX
 PR 24-APR-2002; 2002US-00131831.
 XX
 PR 20-DEC-2002; 2002US-00325899.
 XX
 PA (EXPR-) EXPRESSION DIAGNOSTICS INC.
 XX
 PI Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M,
 PI Rosenberg S;

XX
 DR WPI: 2004-400724/37.
 XX
 PT Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
 PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
 PT rejection, in an individual, comprises detecting the expression level of
 PT the genes.
 XX
 PS Claim 80; SEQ ID NO 335; 1762bp; English.
 XX
 CC The present invention relates to diagnosing or monitoring transplant
 CC rejection, e.g. cardiac or kidney transplant rejection, in an individual
 CC comprising detecting the expression level of one or more genes. The
 CC methods, system and kits are useful in diagnosing or monitoring
 CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
 CC islet, lung, bone marrow or stem cell transplant rejection,
 CC xenotransplant rejection or mechanical organ replacement rejection, in an
 CC individual. The method is also useful in assessing the immune status of
 CC an individual. The methods are also useful in diagnosing and monitoring
 CC diseases that involve the immune system, e.g. rheumatoid arthritis,
 CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
 CC viral, bacterial or fungal infection. The present sequence represents a
 CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis
 CC and monitoring of allograft rejection and other disorders.
 XX
 SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
 Query Match 10.2%; Score 172.4; DB 12; Length 2191;
 Best Local Similarity 52.1%; Pred. No. 9e-43;
 Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
 QY 737 TGGAACTTGAAGAAAGAGATGATCTGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 796
 DB 964 TGGGCTGTGTCTGATCTTCAACATCCCTGTGTGGATCGCATGTGTGTGATCA 1023
 QY 797 AGAAGAGCTCTATGTGAACGGGAAACCTGCTTTTGAAGGGCTTTGAAAGACAGAG 856
 DB 1024 AGAGCAAGTTCTCTCAATAGGAAACCTTTTATTTTCAACGTTTCAACAGATGAG 1083
 QY 857 AATTCCTGCTGTGAGGAGGACCTTTTATTCATGATGATGATGATGATGATGATGATGAT 916
 DB 1084 ATGCGAGATCCGAGGAAAGGCTTGAAGCTGGCTGTGTGAAAGATTTCAACCTG 1143
 QY 917 TGAATGATCAACGCAATTTCTTCAAGACCTCTCATATCTTCAAGTGAAGTGGC 976
 DB 1144 TTGCTGTGCTGTGCTCAACGCTTTCCTTACAGCCTTATCCATGAGGAGTGA 1203
 QY 977 TGGATCTTGGCAGACGATCGGAATCTTGTGATGAGAAAGCCCGACAGTGTATCA 1036
 DB 1204 TGCAATGTGATCGCTATGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1263
 QY 1037 CAAGTACCA---CTACATCCCGAGACTCAGAAATAGACAGAAACAATAGAAAGA 1093
 DB 1264 CGTCCCGAGATTCTTCAACAGGTTCTCTGCATACCAATGCAGGTGATGAAAGAG 1323
 QY 1094 TGATGACAGACAGAAACCAATCCCAATGTGATCATGTGAGTGTGCGAAACGACAG 1153
 DB 1324 TGGTGTGAGGACAGAAACCAACCCCGGTGTGATGTGTGTGCGCAACGAGCTG 1383
 QY 1154 AGTCCACCATCCAGACGGGAGGTTTCTTCAAGCCCTTATATGAGTCTCCATGAAA 1213
 DB 1384 CTTCCACCATAGATCTGTGTGCTACTTCTTGAATGTGTATGTGTCTCACCAATCTCT 1443
 QY 1214 TGGATGAAACAGCCCGCTTGTATGATGATGATGATGATGATGATGATGATGATGAT 1273
 DB 1444 TGGACCTCTCCGCTGTGATCTTGTGATGATCTTGTGATGATCTTGTGATGATCTT 1491
 QY 1274 GAGACGTGCGCTGAAATCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 1333
 DB 1492 CAGACAGAGGGGCTCGTATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 1551
 QY 1334 TCTATCAGGAAAGATGAAAGACCTTCAAGCTTGGAAAAACATAGAAAGCTCT 1393

```

Db      1552 ACGACTACGGGCACTGAGTTGATTCAAGCTGAGCTGGCCACCCAGTTTGAAACTGGT 1611
Qy      1394 ATGCAAGGCACAGAAAGCCCATCTTGTACAGAAATTGGTTCGAGCGGATAGCTGGCA 1453
Db      1612 ATAGAAAGTATCAGAAAGCCCATTTATTGAGAGGAGTATGAGCAGAAACGATTTGCAGGGT 1671
Qy      1454 TCCACTACGATCCACTCAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
Db      1672 TTCACCAAGATCCACTCTGATGTTTCACTGAAAGATACCAGAAAAGTCTGCTAGAGCAGT 1731
Qy      1513 ACGATCAGGCTCCTTTGAAAAA----AGACTACATCATCGAAACACACGTGTGGGCTT 1567
Db      1732 ACCATCTGGGCTGGATCAAAAAACGAGAAAATATGTGGTTGAGAGCTCATTTGAAATT 1791
Qy      1568 TTGCAGATTTTAAGACTCTCTCAGAAATGTGAGAAAGCCATTCTCAACCAAGAGGTGTTT 1627
Db      1792 TTGCCGATTTCAATGACTGAACAGTCAACGACGAGAGTGTCTGGGAAATAAAAAGGGGATCT 1851
Qy      1628 TCACAAGAGACAGACAAACCAAACTGTTGCTCATGTACTGAGA 1671
Db      1852 TCACTCGGCAGAGACAAACCAAAAGTGCAGCGTTCCTTTGGCA 1895

```

Search completed: September 1, 2005, 22:25:44
 Job time : 985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 21:07:30 ; Search time 5837 Seconds
(without alignments)
11014.314 Million cell updates/sec

Title: US-09-936-759-14

Sequence: 1 atggtcaagacgcgaacgaacaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: gb_esc1: *
2: gb_esc2: *
3: gb_hic: *
4: gb_esc3: *
5: gb_esc4: *
6: gb_esc5: *
7: gb_esc6: *
8: gb_g881: *
9: gb_g882: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174	10.3	1377	3	CR593823 full-length
2	174	10.3	1853	6	CD014093 90135266
3	174	10.3	1995	6	CD014092 90134967
4	161.6	9.6	1124	6	CD503076 CDA60-C07
5	161.2	9.5	1138	6	CD503098 CDA60-D07
6	159.6	9.4	2274	3	AK041058 Mus muscu
7	150.6	8.9	816	7	CK467358
8	144.4	8.5	906	5	BQ941196 AGENCOURT
9	142.6	8.4	878	7	CN155220 942501 MA
10	141.6	8.4	1051	5	BX363460 BX363460
11	132.6	7.9	767	7	CF255373 mdv127 C
12	131	7.8	846	6	CA453907 AGENCOURT
13	129.8	7.7	900	6	CA489317 AGENCOURT
14	129.8	7.7	914	6	CB203472 AGENCOURT
15	128.2	7.6	862	7	CO394571 AGENCOURT
16	126.4	7.5	2473	3	AY321342 Ratfus no
17	121	7.2	857	5	BX745933 BX745933
18	117.4	7.0	1287	6	CD503099 CDA60-D07
19	117.2	6.9	802	5	BU357212
20	116	6.9	725	5	BU315768
21	114.2	6.8	724	5	BM441165 BM441165
22	112.6	6.7	771	4	BG121498 602352830
23	111.6	6.6	790	7	CO559438 AGENCOURT
24	110.8	6.6	740	7	CF521612 AGENCOURT

25	109.6	6.5	752	4	BU733141 BU733141
26	109.4	6.5	1268	6	CD503077 CDA60-C07
27	107.6	6.4	673	6	CB841568 M15E-2121
28	107.4	6.4	791	5	BU239978 60323761
29	107.2	6.3	915	5	BQ678153 AGENCOURT
30	106.6	6.3	543	5	BQ360163 OVO-07003
31	106.6	6.3	615	4	BG568800 602588157
32	106.4	6.3	975	4	BG285429 602406582
33	105.6	6.3	616	2	BE179118 RC0-HT061
34	105.6	6.3	637	1	AJ724809 AJ724809
35	105.4	6.2	504	5	BU917201 EST143 BO
36	105.2	6.2	736	5	BP454056 BP454056
37	104.6	6.2	757	5	BI851705 603377940
38	104	6.2	877	7	CF406520 CH3#042_G
39	101.8	6.0	1023	5	BX363459 BX363459
40	101.6	6.0	773	7	CF748482 UI-M-HU-0
41	101	6.0	777	4	BG741172 602631852
42	100.8	6.0	667	7	CN264045 170004243
43	100.6	6.0	548	2	BE751040 202745 MA
44	100.2	5.9	799	5	BU366987 603585124
45	100	5.9	701	5	BX611242 BX611242

ALIGNMENTS

RESULT 1
LOCUS CR593823 1377 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DL001YM21 of B cells (Ramos cell line)
Cot 25-normalized of Homo sapiens (human).
ACCESSION CR593823
VERSION CR593823.1 GI:50474630
KEYWORDS HTC; CNSLT; CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1377)
Li W.B., Gruber C., Jessee J., and Polayer D.
Full-length cDNA libraries and normalization
Unpublished
Contract : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paraday Avenue
Genoscope.
2 (bases 1 to 1377)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source location/Qualifiers
1. 1377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL001YM21"
/tissue_type="B cells (Ramos cell line)" Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 10.3%; Score 174; DB 3; Length 1377;
Best Local Similarity 52.2%; Pred. No. 7.7e-40;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY 737 TGGAGCTTGAAGAAACAGTACTCTGAGCATCGGATCGAGCATGCTGCGAGCG 796
DB 215 TGGAGCTTGTGCTGACTTACACATCTCCCTGTGGGATTCGACGTGCGTGCACCA 274

OY	797	AGAAAGGCTCATATGGAACGGGAAACCGTCTTTTGAAGGCGTTTGGAAAGCAGAG	856
Db	215	AGAGCAATTCTCATCAATGGGAAACCTTTTATTTCCAGGTGTCAACAGCATAGAG	334
OY	857	AATTCGCCGTTCTGGGGCAGGGCACCTTTTATTCATTTGATGATAAAGACTTCACTTC	916
Db	335	ATGCGGAGATCCGAGGGAAAGGGCTTCGACTGGCCGCTGGTGAAGGACTTCAACTGCG	394
OY	917	TGAAGTGATCAAGCGGAATTCCTTCAGAGCCTCTCATCTATCTTACAGTGAAGTGGC	976
Db	395	TTGCTGGCTTGGTGTCCAAACGTTTCCGATCAGCAGCACTCCCTATGCAAGGAAGTGA	454
OY	977	TGATCTTGGCCGACAGACTCGGAATCTTGATGATAGAGAAAGCCCCGACGTTGGTATCA	1038
Db	455	TGCAATGTGTGACCGGCTATGGGATTTGGTGTATGATGAGTGTCCGGCGTGGGCTGG	514
OY	1037	CAAGGTACCA--CTACAATCCGAGACTCAGAGATATGACAGAGACAACATTAAGAGA	1097
Db	515	CGTGCCGCGAGTTCTTTCACAAACGTTTCTGTGCATCACCACTGACAGGTGATGGAAGA	574
OY	1094	TGATCGACAGACACAAGAACATCCACAGTGTATGATGGAGTGGGGAACGAAACAG	1155
Db	575	TGTGTGCTAGAGGACAAGAACACCCCGGCTGTATGTGGTCTGTGGCCAAACGACCTG	634
OY	1154	AGTCCAACTCCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCAATGAAA	1213
Db	635	CGTCCCACTTGAATCTGTGCTACTACTTGAAATGGTGTATCGCTCACAAATCCT	694
OY	1214	TGGATCGAAACAGCCCGGTGTGCATGATGAGGATGATGAGCGCACACAGACGAAACAA	1273
Db	695	TGACCCCTCCCGGCTGTGACTTTTGTGAGCACTCTTAATATGACAG-----	742
OY	1274	GAGACGTGCGCGTGAAGTACTTTCGACATCGTCTGTGTGACAAGGTACTACGGCTGTACA	1333
Db	743	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGTACTACTTGTGTATC	802
OY	1334	TCTATCAGAGGAAGATAGAGAAGACTTCAAGCTCTGGAATAAAGACATAGAAAGCTCT	1393
Db	803	ACGATTAAGGGGACCTGGAGTTGATTACGTGACGTGACGACCGCACCCAGTTTGAATCGGT	862
OY	1394	ATGCAGGACACAGAAGACCATCTTTGTGCAACGAATTCGATGCGGACGATAGCTGGCA	1453
Db	863	ATTAAGAAATTATCAGAAGCCCATTTATTCAGAGGATATAGAGACAAAACGATTCAGAGGT	922
OY	1454	TCGACTACGATCCACTCAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAG-	1512
Db	923	TTTACACAGATTCACCTCTGATGTTCACTGAAAGATACAGAAAAGTGTGCTTAGAGCAGT	982
OY	1513	ACGATCAGAGCTCTTTTGAAGAA-----AGACTATCATGAGGAACACAGCGTGGGCT	1566
Db	983	ACCATCTGGGCTTGGATCAAAAACGACAGAAATACGTGGTTGGAAGCTCATTTTGAATT	1042
OY	1568	TTGCAGATTTTAAAGCTCCTCAGATGTGAGAAAGCCATTTCTCAACACAAGGGTGT	1627
Db	1043	TTGCGGATTTTATGACTGACAGTACACGAGCAGAGGTGCTGGGGAATTAAGGGGATCT	1102
OY	1628	TCACAAGACAGACAACCCAAATCTGTGCTCATGTACTGAGA 1671	
Db	1103	TCACTCGGACAGACACAAAAGTGCAGCGCTCTCTTTTGGGA 1146	

RESULT 2	CD014093	1853 bp	mRNA	linear	EST 21-OCT-2003
LOCUS	CD014093				
DEFINITION	90135266	Single gene	library	Homo sapiens	cdna, mRNA sequence.
ACCESSION	CD014093				
VERSION	CD014093.1	GI:37777622			
KEYWORDS	EST.				
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				

REFERENCE	1 (bases 1 to 1853)
AUTHORS	Jin, P., Fu, G. K., Wilson, A. D., Yang, J., Chien, D., Hawkins, P. R., Au-Yang, J., and Stuve, U. L.
TITLE	PCR amplification and cloning of novel splice variant mRNAs from known drug target genes
JOURNAL	Genomics 83 (4), 566-571 (2004)
COMMENT	Contact: Jin, P. Incyte Corporation 3160 Porter Drive, Palo Alto, CA 94304, USA Tel: 650 621 8639 Fax: 650 621 8965 Email: pj@incyte.com.

```

FEATURES
    source
        1. 1853
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone_lib="Single gene library"
                /note="Vector: pDrive Cloning Vector; RT-PCR was performed
                    using gene-specific primers flanking the open-reading
                    frame. PCR products were subcloned into pDrive Cloning
                    vector and sequenced completely using M13 forward and
                    reverse primers. Sequencing gaps were closed by
                    re-sequencing using primers flanking the gapped areas."

```

ORIGIN

Query Match	10.3%	Score 174;	DB 6;	Length 1853;
Best Local Similarity	52.2%;	Pred. No. 8.5e-40;		
Matches 493; Conservative	0;	Mismatches 430;	Indels 21;	Gaps 4

QY	737	TGGAACTTGA	AAAAAGACAGAT	CACTCTGGACAT	TGGAAATCAGAA	CGATCAGTGGAG	CG	796		
Db	624	TGGGGCCCTG	ATCTGACTTTC	AACAACAC	CTCCCTGTGGGGAT	CCGCACTGTGCTGT	CAACA	683		
QY	797	AGAAAGGCT	CTATATCTGA	ACGGGAAAC	CTGTCTTTTGA	GGGCTTTGGAAAG	CA	856		
Db	664	AGAGCGA	TTCTCATCA	TATGGAAAC	CTTTCTATTTT	CCACGGTGTCA	AAAGCAT	743		
QY	857	AATTCCCG	TTCTGTGGG	CAGGGCAC	CTTTTAT	TCCATTGAT	ATAAAG	CTTCAACTTC	916	
Db	744	ATGCGG	AATCCGAGG	GAAGGCTTTC	GCATCGGCTCT	GTGTGA	AGGACTTTC	CAACTTC	803	
QY	917	TGAAGTGA	TCAACGCGA	ATTCTTTCA	GAGACCTCT	CACTATCTT	CTTACAGTGA	AGATGGC	976	
Db	804	TTCCGTGG	CTTTGTGTCC	CAACGCTTTC	CTTACAGC	CACTACCTT	ATGACAGG	GAATGA	863	
QY	977	TGATCTT	TGCGCAG	ACTCGGAA	TCTTGTGAT	GAGAAAG	CCCCGAC	CTTGTGATCA	1036	
Db	864	TGCAGAT	GTGTGACCG	GTATGGAA	TGTGTGCAT	TGATGAG	GTGTCCGGCGT	GGGCTGTG	923	
QY	1037	CAAGTAC	CA--CTACA	TATCCCGAG	ACTCAGAA	GTATG	CAAGACA	CAACTT	1092	
Db	924	CGTGTCCG	CAATTTTCA	CAACGTTTCT	CTGCAT	CAACAATG	CAAGTGA	TGAAGAAG	983	
QY	1094	TGATCGA	CAGACAC	AAAGAAC	CTCCAC	GTGTGAT	GATGTGAG	TTGGGAA	CGAACG	1152
Db	984	TGTGTG	CGATGGGA	CAAGAAC	CCCCGGGT	GTGTGAT	TGTGTGTG	GGCAAG	CGCTG	1042
QY	1154	AGTCCAA	CCATCCAG	ACGGAG	GTTCCTTCA	AAAGCCCTT	TATGAG	CTGCAAT	TGAAA	1213
Db	1044	CGTCCCA	CTTGAATCT	CTGGCTT	ACTTGA	AGATGTG	ATCGCTCA	CAACCAAT	CTCT	1103
QY	1214	TGATTCGA	CAACG	CCCCGTTG	TCAATG	TGAGAC	ATGATGGA	CGCAC	ACAGAGAA	1277
Db	1104	TGAGAC	CTCCCGG	CTGTGAC	CTTTGTGAG	CACTT	TAACATG	CAG-----		1155
QY	1274	GAGAC	GTGGC	CTGAGT	ACTTTCG	ACATCGT	CTGTGTA	ACAGT	ACTACGGCTGT	1332
Db	1152	CAGAC	CAAGGGG	GCTCCG	TATGTGAT	GTGATCTG	TTTGA	CAAGT	ACTACTTGTG	1211
QY	1334	TCTAT	CAGGGA	AGGAT	TAGA	AGGACTTCA	AGCTCTG	GA	AAAAAGAC	1392
Db	1212	ACGACT	ACGGG	CACCTG	AGTTGAT	TTCAGCTG	CAGCTGG	CCACCACTG	TTTGA	1277

QY 1394 ATGCAAGCAGAGAAGCCCATCTTGTGACAGATTGGTGGAGGATAGTGCA 1453
 DB 1272 ATBAGAAGTATCAGAAAGCCATTATTCAGAGCGATATGAGAGAAACATTTGACGGGT 1331
 QY 1454 TCCACTACGATCCACCTCAATGTTCTCCGAGAGTACCAAGCAGAGCTGTTGAAAG- 1512
 DB 1332 TTCACAGAGATCCACTCTGATGTTTCACTGMAAGATACCAAGAAATCTGCTAGACAGT 1391
 QY 1513 AGCATAGGCTCTTTTGAAGAA-----AGACTACATCATGGAACACACGTGGGCGCT 1567
 DB 1392 ACCATCTGGGTCTGATCAAAAACGAGAAATACGTGGTGGAGAGCTCATTTGGAAAT 1451
 QY 1568 TTGCAGATTTTAAGACTCCCTCAGATGTGAGAAATCCCATTTCCACCAAGAGGTGTT 1627
 DB 1452 TTGCCGATTTTCATGACTGAAACATCACCGACGAGAGTCTGGGGAATTAAGGGGATCT 1511
 QY 1628 TCACAGAGACAGACAAACCCAACTGTTGCTCATGTACTGAGA 1671
 DB 1512 TCACCTGGCAGAGACAAACCAAAAGTGCAGCGTTCTTTTGGCA 1555

RESULT 3
 LOCUS CD014092 1995 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90134967 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014092
 VERSION CD014092.1 GI:37777621
 KEYWORDS EST.

SOURCE
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS Buiarocia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1995)
 Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R., Au-Young, J., and Stuve, L.L.

TITLE PCR isolation and cloning of novel splice variant mRNA from known drug target genes

JOURNAL
 COMMENT Genomics 83 (4), 566-571 (2004)

CONTACT: Jin, P.
 Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639
 Fax: 650 621 8965
 Email: pjin@incyte.com.

FEATURES
 Location/Qualifiers

1..1995
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."

ORIGIN

Query Match 10.3%; Score 174; DB 6; Length 1995;
 Best Local Similarity 52.2%; Pred. No. 8.8e-40;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TCGAATTGAAAGAGAGTCACTCTTGAGATCGGAATCGAAGCATCAGCTGGAGG 796
 DB 766 TGGGGCTGTGTCTGCTCTACACACTCTGTGGGATCGCACTGTGCTGTACCA 825
 QY 797 AAGAAGGCTTATCTGAACCGGAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
 DB 826 AAGACGAGTCTCTCATCAATGGAAACCTTTTATTTCCAGGTGCAACAGCATGAG 885
 QY 857 AATTCCTGCTTGGGGCAGGACCTTTATCATGATGATTAAGACTCAACCTTC 916

DB 886 ATCCGACATCCGAGGGAAGGGGCTTCGACTGGCCGCTGTGGTGAAGACTTCAACCTGC 945
 QY 917 TGAATGATTCACCGGAATTCTTTTCAGACCTCTCATCTATCTTTCAGATGAAGTGC 976
 DB 946 TTCGTGCTGTGGTGTGCAACGGCTTTCCTGACAGCCACTACCTCTATGCAAGAGTA 1005
 QY 977 TGGATCTTCCGACAGACTCGGAATCTGTGATGAGAAAGCCCGCAGTTGATCA 1036
 DB 1006 TGCAGATGTGTACCGCTATGAGATGTGTATCATCATATGATGTCTCCGCGTGGGCTGG 1065
 QY 1037 CAAGTACCA---CTACAAATCCCGAGACTCAGAAAGATACAGAAAGCAATTAAGAA 1093
 DB 1066 CGTCCGCGAGTTCTTCAACAAAGTTCTCTGATATACCAATGACAGATGAGGAAG 1125
 QY 1094 TGAATGACAGACAAAGAACATCCCATGTGTATCATGTGAGTGTGGGAAGCAACAG 1153
 DB 1126 TGTGCTGTGGGACAAAGAACACCCCGCGTGTGTGTGTCTGTGGGCAAGACCTG 1185
 QY 1154 AGTCCAAACCATCCAGACGCGGAGGGTCTTCAAAAGCCCTTATGAGACTGCAATGAA 1213
 DB 1186 CGTCCCATTAAGATCTGTGCTACTACTTGAAGATGTGATGTGCTACACCAAAATCCT 1245
 QY 1214 TGAATGAAACAGCCCGCTTGTATGATGATGATGAGATGAGACGACCAAGAGAA 1273
 DB 1246 TGGACCCCTCCGGGCTGTGACCTTTGTGAGCACTTATATGAG----- 1293
 QY 1274 GAGAGTGGCGCTGAAGTACTTGCATGCTGTGTGTGAACAGTACTACGCTGTGTACA 1333
 DB 1294 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGTACTCTTGGTATTC 1353
 QY 1334 TCTATCAGGGAAGATAGAAAGAGACCTTCAAGCTCTGGAAGAAACATTAAGAGCTCT 1393
 DB 1354 AGACTACGGGACCTGAGGTGATTAAGTCAAGTCAAGTGCACCCAGTTTGAAGTGTGT 1413
 QY 1394 ATGCAAGCAGAGAAAGCCCATCTTGTGTACAGAAATTCGATGCGAGCGAGATAGTGCA 1453
 DB 1414 ATBAGAAGTATCAGAAAGCCATTATTCAGAGGAGATATGAGACAGAAAGATTTGACGGGT 1473
 QY 1454 TCCACTACGATCCACTCAATGTTCTCCGAGAGTACCAAGACAGACTGTTGAAAG- 1512
 DB 1474 TTCACAGAGATCCACTCTGATGTTTCACTGAAAGATACAGAAAGTCTGTAGACAGT 1533
 QY 1513 AGCATCAGGCTCTTTTGAAGAA-----AGACTACATCATCCGGAACACACGTGGGCGCT 1567
 DB 1534 ACCATCTGGGTCTGATCAAAAACGAGAAATACGTGTGAGAGCTCATTTTGGATTT 1593
 QY 1568 TTGCAGATTTTAAGACTCCTCAGATGTGAGAAAGCCCATTTCCAAACCAAGAGGTGTT 1627
 DB 1594 TTGCCGATTTTATGACTGAACAGTCAACGACGAGAGTGTGGGAATTAAGGGGATCT 1653
 QY 1628 TCACAGAGACAGACAAACCCAACTGTTCTCATGTACTGAGA 1671
 DB 1654 TCACCTGGCAGAGACAAACCAAAAGTGCAGCGTTCTTTTGGCA 1697

RESULT 4
 LOCUS CD503076 1124 bp mRNA linear EST 12-JUN-2003
 DEFINITION CDA60-C07.x1d-t-SHG-CDA Gasterosteus aculeatus cDNA clone
 ACCESSION CD503076
 VERSION CD503076.1 GI:31432977
 KEYWORDS EST.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus
 Buiarocia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphae; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.
 1 (bases 1 to 1124)
 Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M.
 Expressed sequence tags from Gasterosteus aculeatus

preceded by a synthetic XhoI site. 5 prime adaptor were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of PBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual PBK-CMV phagemid clones for EST sequencing."

ORIGIN

Query Match 9.5%; Score 161.2; DB 6; Length 1138;
Best Local Similarity 51.9%; Pred. No. 4,7e-36;

Matches 473; Conservative 0; Mismatches 418; Indels 21; Gaps 4;

```

Oy 748 AAGAGCAGTACACTCTGACATCGGAATCAGAACGATGCTGGAGCAGAGAGAGGCTC 807
Db 71 AAGCAGCTGTACACTTACCCAGTCGGCATCCGACGGTGCAGCTTACGACACCCAGTTTC 130
Oy 808 TATCGAAGCGGAAACCTGCTTTTGGAAAGGCTTTGGAAAGCAGAGAAATTCCTCCGTT 867
Db 131 CTCATCAACAAAAGCCCTTCTTACTTCCACGAGTAATTAACACGAGACTCTGATATT 190
Oy 868 CTGGGCGCAGGCGACCTTTTATCCATGATGATTAAGACTTCAACCTTCTGAAGTGATC 927
Db 191 CGAGGCAAGGCTGTGACTGGCCCTCATGTGAGAGACTTTTAACTTATGAAGTGGTTG 250
Oy 928 AAGCGAATTTCTTTCAGACCTTCTCACTATCTTACAGTAAGAGAGTGCGTGAATCTTGC 987
Db 251 GGGGCCAATCTGTTCCGACACAGCCACTACCTTATGACAGAGAGATCTTGAGATGTGT 310
Oy 988 GACAGACTGGGAATCTTGTGATAGCGAAGCCCGCAGCTTGTATCA---AGGTAC 1044
Db 311 GACCGCGATGGAGCTGTGTGATAGCGAGTCCCGGCGTGGCATTAAGACATCTCGC 370
Oy 1045 CACTACAAATCCCGAGACTCAGAAAGATAGCAGAAACAAATTAAGAAATGATTCAGACA 1104
Db 371 AATTTTGGAAAGCCCTCTTAAACCATCACTGTGTGTATGAGCAGAGCTGTACGTCCG 430
Oy 1105 CACAAGAACCATCCAGTGTGATGATGAGAGTGGCGAAGCAAGCAGAGTCCAACT 1164
Db 431 GACAAGAACCATCCCTGTGTGATGATGAGTGGCGCAATGAGCCGCGTCAGAGATG 490
Oy 1165 CGAAGCGCGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGATGGAACA 1224
Db 491 CCTCCTGCTGATTAATTTCAAAACCTTGATTAACAAATTAAGATTCACAAAC 550
Oy 1225 CGCCCGTGTGATGATGAGATGAGCAGCAGCAGAGAGAAACAAGACGCTGGCG 1284
Db 551 CGGCCCGT-----CACTTTTATCAGACAGACATTAATCTAGTCCAGGATTAAGGG 598
Oy 1285 CTGAAGTACTTGCAGATCGTCTGTGTGAACAGTACTACGCGTGTATCTATCAGGGA 1344
Db 599 GCTCCCTAGCTGGAAGTGTATGCGGTAACAGTTTCTCCTGATACATAGCCCGGCG 658
Oy 1345 AGATAGGAAGAGACTTCAAGCTCTGGAAAAAGACATGAAAGAGCTTATGCAAGCGAC 1404
Db 659 CACCGGAGGTCACTCCCATCCAGCTCAACACTGATTGAAAGATGTGACGAAAGTAC 718
Oy 1405 AGAAAGCCCATCTTTGTCAAGAAATTCGGTGGGAGCGGATGAGTGGCATTCACATACAT 1464
Db 719 CAGAAACCATATCCAGAGCGAATTCGAGAGGAGATGCGGTGCGGGCTTCAAGTGTAT 778
Oy 1465 CCACCTCAAAATGTTCTCCGAGAAGTACCAAGCAGAGCTGTTGAAAGAAGCATCAGGCTC 1524
Db 779 CCACCGGTATGTTTACTGAGAGGTACGAAAGTTATGTCCTGAGAGGTACACACAACGTC 838
Oy 1525 CTTTGGAAAAAG-----ACTACATATATGGAAACACAGCTGTGGCGCTTTGCAATTTTA 1579
Db 839 TTTCAGCAAGAAAGAGCAGTACGTATCGGGAACTCATCTGGAACCTTTGACAGACTTCA 898
Oy 1580 AGACTCTCAGAAATGTGAGAAGACCATCTCAACCAAGAGGAGTGTTCACAAAGAGACA 1639

```

```

Db 899 TGACCAACCAAGGATCATGCG-CTGTGTGGGGAACAAGAGGCTGTCTTACACAGGAAA 957
Oy 1640 GACAACCCCAAC 1651
Db 958 GGCAGCCCAAGC 969

RESULT 6
AK041058
LOCUS
DEFINITION
Mus musculus adult male aorta and vein cDNA, RIKEN full-length
structural library, clone:A530072005 product:beta-glucuronidase
AK041058
VERSION
AK041058.1 GI:26088231
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
99279253
PUBMED
10349636
REFERENCE
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
2049374
PUBMED
11042159
REFERENCE
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Teshiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
20530913
PUBMED
11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
5
PUBMED
11076861
TITLE
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
6 (bases 1 to 2274)
PUBMED
11076861
REFERENCE
Adechi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
TITLE
Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of
JOURNAL

```

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse cDNAs.

Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

Location/Qualifiers

1..2274
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/sex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14..1778
/note="beta-glucuronidase structural (MGD|MG1:95874, GB|NM_010368, evidence: BLASTN, 99%, match=733)
putative"
2260..2265
polya_signal
/note="putative"
2274
polya_site
/note="putative"

ORIGIN

Query Match 9.4%; Score 159.6; DB 3; Length 2274;
Best Local Similarity 50.6%; Pred. No. 1.8e-35;
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

739 GAACCTTGAAGAAAGACGATGACCTGACATCGCAATGACACGATCAGCTGGAGCAG 798
759 GAGCTGTGACGATGACATGACACCTCTCTATCGGATTCGAAACGTGCTGTCAAAAG 818
799 AAGAGCTCTATCTGAAAGGAAACCTGCTTTTGAAGGGCTTGAAGACGAGAGA 858
819 AACAAAGTCTCATTAACGGAAGCCCTTCTATTTCAGGGGTCAATAGACAGAGAT 878
859 TTCCTCCGTTCTGGGGGACGACCTTTATCCATTGATGATAAAGACTTCAACCTTCTG 918
879 TCAGATATCCGAGGGGAAAGGCTTCGACCTGGCGCTGTGTAAGGATTTCAACCTGCTC 938
919 AAGTGATCAAGCGCAATTTCTTTCAGGACCTCTCACTATCTTTCAGTGAAGTGGCTG 978
939 CGTTGGCTCGGGGCAAAATTCCTTTCCTACACACTATCCCTACTCAAGAGGATCACTT 998
979 GATCTTGGCAGACGACTCGGAATCTCTTGATAGACGAAGCCCGACGTTGG3---TATC 1035
999 CAGCTCTGACGATCGGAGATGTGATCATGATGATGTCCTGGTGTGGGCAATTGTG 1058
1036 ACAAGTACCACTACATCCGAGACTCGAAGATAGCAGAAAGACAACATGAAGAAATG 1095
1059 CTACCTCAGAGTTTGGCAACGAGTCACTTCGGCACCACTAGAGGTGATGAGAGGCTG 1118
1096 ATCGACAGACAAAGAACATCCCACTGATGATGATGATGATGATGATGATGATGATG 1155
1119 GTTCCGCGGAGCAAAATCACTCGGCTGTGATGATGATGATGATGATGATGATGATG 1178
1156 TCCAAACATCCAGAGCGGAGGTTCTTCAAGCCCTTATGAGATGATGATGATGATGATG 1215
1179 TTCTGCTTAAACCCGCGCATATTAATCTTAAAGCGCTATATCAACCAACCAAGCCCTG 1238
1216 GATCGAACAGCCCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1275

1239 GACCTCACCCGTCCTGATGACCTTTGTGACA-----ACGCCAAATATGATGCA 1286
1276 GACGTGGGCTGAAAGTACTTGCATGATGATGATGATGATGATGATGATGATGATG 1335
1287 GACCTGGGGGGCCCGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1346
1336 TATCAGGAGAGATAGAAAGAGACTTCAAGCTCTGAAAAAGACATGAGAGCTTAT 1395
1347 GACATGAGGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1406
1396 GCAGGACACAGAAAGCCATCTTTGTGCAAGAAATTCGGTGGGACGATGATGAGCATC 1455
1407 AAGAGCATCAGAAAGCCATTTATCCAGAGCAGATGAGAGACGACATCCAGAGATC 1466
1456 CACTACGATTCACCTCAATGTTTCTCCGAAGATGACCAAGAGAGCTGTTGAAAGACG 1515
1467 CACGAGACCCCGCTCGATGTTGATGAGAGATGACCAAGAGGCTGTTCTGAGAAATTAC 1526
1516 ATCAGGCTCTTTTGAAG-----AAGACTATCATCATCGGAACACAGCTGTGGCTTT 1569
1527 CATTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1586
1570 GCAATTTTAAAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1629
1587 GCCGACTTATGACGAAACGATGATGATGATGATGATGATGATGATGATGATGATG 1646
1630 ACAAGAGACAGAACCAACCACTGTTGCTCATGATGATGATGATGATGATGATGATG 1671
1647 ACTGCGCAGAGACAGCCCAAACTTGGCTTTTATTTTGGCA 1688

RESULT 7

LOCUS CK467358 816 bp mRNA linear EST 14-JAN-2004
DEFINITION 938649 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION CK467358
VERSION CK467358.1 GI:40838635
KEYWORDS EST.

SOURCE

Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE

1 (bases 1 to 816)
Smith, T.P.L., Fekking, B.A., Ford, J.J., Vallet, J.L., Wise, T.A.,
Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
Unpublished (2003)

JOURNAL

CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smtl@email.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alc option. Vector identified with
cross_match v0.990329.
Plate: TM8042 row: M column: 18
Seq primer: GTAATGACATCCTACTATAGGG.

FEATURES

source

1..816
location/Qualifiers
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

ORIGIN

Query Match 8.9%; Score 150.6; DB 7; Length 816;
 Best Local Similarity 53.2%; Pred. No. 6,1e-33;
 Matches 396; Conservative 0; Mismatches 334; Indels 15; Gaps 3;

QY 751 GACGAGTACACTCTGGACATCGGAATCAGAACGATGAGTGGACGAGAAAGGCTCTAT 810
 DB 84 GACTTCTACACCTCCCTGGGGATTTCACACGCTGTGTCAAGAGAGACAGTTCTTC 143
 QY 811 CTGAACGGGAAACCTCTCTTTTGAAGGCTTTGAAACGAGAGAAATTCCTCCGTTCTG 870
 DB 144 ATCAACGGGAAACCTTTCTATTCTTCATGGGGTTTACACACAGAGATCAGATATCCCA 203
 QY 871 GGGCAGGAGCCTTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 930
 DB 204 GCGAAGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 263
 QY 931 GCGAATCTTTTGAAGCCTCTCACTATCTCTTACAGTGAAGTGGCTGATCTTTCCGAC 990
 DB 264 GCGAAGCCTCTCGACACGACCTACCTACCTACGAGAGAGATGATGATGATGATGATGAT 323
 QY 991 AGACTCGGAATCTCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1047
 DB 324 CATTATGGGATCGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 383
 QY 1048 TACAATCCCGAGACTCAGAAAGATAGCAGAAACAATAGAAAGATGATGATGATGATGAT 1107
 DB 384 TTCAAGACGCTCTCTGACGACCACTAGAGGTGATGATGATGATGATGATGATGATGAT 443
 QY 1108 AAGAACCATCCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167
 DB 444 AAGAACCAACCGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 503
 QY 1168 GACGCGGAGGCTTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGATGATGATGAT 1227
 DB 504 CAGGCTGCTTACTACTCAAGATGCTGATGCTGACACCAAAAGCCCTTGACCCCTCCG 563
 QY 1228 CCCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
 DB 564 CCCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611
 QY 1288 AAGTACTTGTGACATGCTGTGTGTGAACAGATCTACGCTGTGATCTATCTATGAG 1347
 DB 612 CCGTATGTGACATGATCTGTGTGTGAACAGATCTACTCTGTGTATCATGATCAAG 671
 QY 1348 ATGAAGAAGAGCTTCAAGCTCTGGAAGAAAGACATGAAAGCTCTATGCAAGGAC 1407
 DB 672 ATGAAGTGTATCAAGTGTGACGCTGCGCACCCAGTTGAGAGGTGATGAGGCTTAC 731
 QY 1408 AAGCCATCTTGTGACAGAAATTCGTTGGGACGCGATGATGATGATGATGATGATGAT 1467
 DB 732 AAGCCATTTATCAAGAGCGATGACGAGGAGAAACCATATAGGTTTACAGAGATCCA 791
 QY 1468 CCTCAATGTTCTCCGAGAGATACC 1492
 DB 792 CCTCATGTTTCAGTGAAGAAATACC 816

RESULT 8
 BQ941196 906 bp mRNA linear EST 21-AUG-2002
 LOCUS BQ941196 AGSCOURT 8741562 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6420519
 DEFINITION 5' mRNA Sequence.
 ACCESSION BQ941196
 VERSION BQ941196.1 GI:22356674
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 906)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/BTP/Gazdar
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2594 row: 9 column: 16
 High quality sequence stop: 765.
 Location/Qualifiers
 1. 906
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6420519"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_11b="NIH_MGC_18"
 /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN
 Query Match 8.5%; Score 144.4; DB 5; Length 906;
 Best Local Similarity 52.1%; Pred. No. 4,5e-31;
 Matches 454; Conservative 0; Mismatches 396; Indels 22; Gaps 5;

QY 828 CTTTGTGAAGGCTTTGGAAGACGAGAAATTCCTCTGAGGAGGAGCCTTTTA 887
 DB 1 CTAATTTCAAGGTGTCAACAGCATGAGATCGGACATCGAGGAGGCTTGACATG 60
 QY 888 TCCATTGATGATTAAGAACTTCAACCTTGTGAATGATCAAGCGAATCTTTTACAGAC 947
 DB 61 GCGGCTGCTGGGAAAGGACTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 QY 948 CTTCATGATCTCTTACATGATGAAAGTGGCTGTGATCTTCCGACACATCGGAATCTTGT 1007
 DB 121 CAGCCACTACCTCTATGACAGAGAAAGTGAATGTGTGACCGCTATGGGATTTGTGT 180
 QY 1008 GATAGACAAAGCCCGCAGCTGGTGTATCAAGATGACCA---CTACAAATCCGAGACTCA 1064
 DB 181 CATCATGATAGTCTCCGCGCTGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QY 1065 GAAGATGACAGAAACAACTAAGAAAGATGATGACAGACACAAAGAAACCATCCAGTGT 1124
 DB 241 GCATCACCAATGACAGTGTATGAGAAAGTGTGTGTGAGGACAAAGAACACCCGCGGT 300
 QY 1125 GATCATGTGAGATGTGGGAAAGCAACAGATCCCAACATCCAGACGGGAGGTTTCTT 1184
 DB 301 CGTATGTGTGTGTGGCCAAAGAGCTGTGCTCCCACTGAATGTGTGTGTACTACTT 360
 QY 1185 CAAGCCCTTTATGAGACTGCAATGAATGATGCAACACGCGCTGTGTGATGATGATGAT 1244
 DB 361 GAAGATGTGATGTGCTACACCAATCTTTGAGACCTTCCGCTGTGTGATGATGATGAT 420
 QY 1245 CATGATGACGACACAGACGAGAGAAAGAGAGAGCTGCGCTGTAAGTACTTGACATCGT 1304
 DB 421 CAATCTTAATGATGAG-----CAGACAAAGGGGCTCGTATGTGATGATGAT 468
 QY 1305 CTGTGTGAAGAGTCTACGCTGTGTATCATCTATGAGGAGAGATGAGAAAGAGACTTCA 1364
 DB 469 CTGTTTGAACAGTACTCTCTTGTATCAGACTACGAGGACCTGTGAGTTGATCACT 528
 QY 1365 AGCTGTGAAAAAGACATAGAAAGAGCTTATGCAAGGACAGAAAGCCCATCTTTGTGAC 1424
 DB 529 GCAGCTGGCCACCCAGTTTGAAGACTGTATTAAGAAAGTATCAGAAAGCCCATTTATCAG 588


```

QY      1108 AAGAACCATCCGAGTGTGATCATGTGAGTGTGGCGAAGCAACGAGTCCAAACCATCA 1167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      362 AAGAACGAGCCCTTCATGTGTGATGTGTGATGAGTACGATGAGCCGAGCATCAGAGCTGCC 421
QY      1168 GACGGCGAGGGTCTTCTTCAAGCCCTTTATGAGACTGCCAATGAAATGATGAACAGC 1227
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      422 CCAGCAGCTTACTACTTCAGACACTGATGATGACACTTAAGACTGTGATCCCTCCAGA 481
QY      1228 CCCCCTGTGATGATGAGCATGATGAGCAGCAGCAGCAGAGAGAAACAAGAGACTGGCGCTG 1287
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      482 CCGTGAACGTTGTGTGTGATCTAATTAACGCTGTC-----GATCATGTGCT 529
QY      1288 AAGTACTTGACATCGTCTGTGTGAACAGGTACTACGCTGTATCATCTATCAGGAAAG 1347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      530 CCTTATGTGATGATGATCTGTGTGAACAGCTACTTCTCTGTATCAGACAGCCAGCCAT 589
QY      1348 ATGAGAAGAGAGACTTCAAGCTGTGAAAAAGACATPAGAAGAGCTGTATGCAAGCAGCA 1407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      590 CTGGAAGTTATTCCACTGCAACTTACAGCAGAGTTGAGAACTGGTATPAAACTTACAG 649
QY      1408 AAGCCATCTTGTGACAGAAATTCGTGGGAGCGGATGCTGTCATCCATCAGATGCA 1467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      650 AAACCATTTATCCAGAGTAATATGAGCTGAGTTCCGAGCTTCACAGTATCCA 709
QY      1468 CCTCAATGTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAAG 1512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      710 CCGTGTATGTTCAGTGAGAGATATCAGAAAGCTATGCTGAAGAG 754

```

```

RESULT 12
LOCUS      CA453907          846 bp      mRNA      linear      EST 12-NOV-2002
DEFINITION AGENCOURT_10738557 MAPCL Homo sapiens cDNA clone IMAGE:6718553 5',
            mRNA sequence.
ACCESSION  CA453907
VERSION     CA453907.1 GI:24903124
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 846)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga9bs-r@mail.nih.gov
            Tissue Procurement: Kristi A. Eglund, Ira Pastan
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LLM14275 row: a column: 17
            High quality sequence stop: 708.
            Location/Qualifiers
                1..846
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6718553"
                /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
                hTERT-HME1, LNCaP"
                /lab_host="EMDH108"
                /clone_1fb="MAPCL"
                /note="Vector: PCWV-SPORT6, Site 1: EcoRV, Site 2: Not I;
                Subtracted with brain, liver, lung, kidney and muscle.
                directionally cloned. Priming method: oligo-dT. Average
                insert size: 1800 bp. Library amplification: 26,000 fold.
                Kristi A. Eglund, James J. Vincent, Robert Strausberg,
                Bungkok Lee & Ira Pastan: Discovery of new breast
                cancer genes encoding membrane and secreted proteins.

```

ORIGIN Manuscript submitted."

Query Match 7.8%; Score 131; DB 6; Length 846;

Best Local Similarity 52.4%; Pred. No. 4,4e-27;

Matches 417; Conservative 0; Mismatches 356; Indels 23; Gaps 5;

```

QY      866 TTCTGGGCGAGGGACCTTTATCATGATGATPAAAAGACTTCAACCTTGTAAGTGA 925
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 TCCAGGAGAGGGCTTGACTGCGCGCTGTGTGAGAGACTTCACCTGCTCGCTGGC 120
QY      926 TCAACGCGAATCTTTCAGAGCTCTCACTATCCCTTACAGTGAAGAGTGGCTGATCTTG 985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 TTGGTGCCACGCTTTCGATACAGCAGCACTACCCCTATGAGAGAGAAATGATGAGATGT 180
QY      986 CCGACAGACTGGAATCTTGTGATAGCAGAGCCCGGACAGTGTGTATCAGAGTACC 1045
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      181 GTGACCGCTATGGGATGTGTGATCATCATGATGTGCCGGGTGGGCTGGCGCTGCCG 240
QY      1046 A---CTACATCCGAGACTCAGAAATAGCAGAAAGCAACATPAAAGAAATGATCGACA 1102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241 AGTTCTTTACACAGCTTTCTGTGATCAGCAGATGAGAGTGAAGAGATGTGCTGA 300
QY      1103 GACACAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACGAGTCCAAAC 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      301 GGGACAAAGAACACCCGCGGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
QY      1163 ATCCAGACGCGGAGGGTCTTTCMAAGCCCTTTATGAGACTGCCAATGAATGATCGAA 1222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      361 TAGAATCTGTGGGTACTACTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY      1223 CACGCGCGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      421 CCGGCTGTGATCCTTTGTGAGCACTTAATATGACG-----CAGACAAAG 468
QY      1283 CGTGAAGTACTTGCATCTGTGTGAAACAGATCTACAGCTGATCATCTATTCACAG 1342
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      469 GGGCTCGGTATGTGATGTGATCTGTTGAAACAGTACTACTCTGTTGATATCAGCATCAG 528
QY      1343 GAAGATGAGAAAGAGACTTACAGCTTGGAAAAAGACATGAAAGACTCTATGCAAGC 1402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      529 GGCACCTGGAATGTATTCAGCTGACGTGGCCACCGATTGAGAACTGTATTAAGAAGT 588
QY      1403 ACAGAAAGCCCATCTTGTGTCACAGAAATTCGTTGGGAGCGGATGCTGCATCCACTAG 1462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      589 ATCGAAGCCCATTTATTCAGAGCGATGTGAGAGCAAGAAACATTTGCAGGGTTTCAACAG 648
QY      1463 ATCCACCTCAATGTCTTCCGAGAGTACCAAGCAGAGCTGTTGAAAAG-ACGATCAG 1521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      649 ATCCACCTGTGATGTTCAGTAAGATACAGAAAAGTCTGCTAGAGCATACATCTGG 708
QY      1522 CTCCTTTGAAAAAAGAC-----TACATCATCGGAACACAGCTGTGGCCCTTTGCAAT 1575
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      709 GNTCTGATCAAAAAACGAGAAATACGTGTGTGAGAGCTCATTTGGAATTTTGCCGAT 768
QY      1576 TTTTAAAGTCTCTCGAATGTGAGAAAGCCATTTCCAA-CCACAAGGGTGTTTTCAAG 1634
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      769 TTCAATGACTGAACGTACCGACAGAGAGTGTGGGAAATTTAAAGGGATCTTCACTCG 828
QY      1635 AGACAGCAACCCCAA 1650
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      829 GCAGAGACCAACCCCAA 844

```

```

RESULT 13
LOCUS      CA489317          900 bp      mRNA      linear      EST 14-NOV-2002
DEFINITION AGENCOURT_10810259 MAPCL Homo sapiens cDNA clone IMAGE:6721700 5',
            mRNA sequence.
ACCESSION  CA489317
VERSION     CA489317.1 GI:24952108
KEYWORDS   EST.
SOURCE      Homo sapiens (human)

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 900)
 NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 cDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
 Cloning Strategy: Agencourt Bioscience Corporation
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 http://image.llnl.gov
 Plate: LLM14283 row: d column: 20
 High quality sequence stop: 580.
 Location/Qualifiers
 1..900
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6721700"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 HERT-HME1, LNCaP"
 /lab_host="EMD108"
 /clone_1ib="MAPCL"
 /note="Vector: PCMV-SPORE; Site 1: EcoRV; Site 2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 directionally cloned. priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bangkok Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN
 Query Match 7.7%; Score 129.8; DB 6; Length 900;
 Best Local Similarity 51.9%; Pred. No. 1e-26;
 Matches 402; Conservative 0; Mismatches 352; Indels 21; Gaps 4;

866 TTCTGGGGGAGGAGCCTTTATTCATGATGATTAAGACCTTTCAGAGTGA 925
 31 TCCGAGGAGGAGGCTTTCAGCTGCGCTCTGTGAAGACCTTTCAGCTGCG 90
 926 TCAAGCGGAGATCTTTCAGAGCCTTCATCTCTTCAAGTGAAGAGTGGATCTTG 985
 91 TTGTGCGCAACGCTTCCCTACAGCACTACCCCTTTCAGAGAGTGAATGATGAT 150
 986 CCGACAGACTCGGAATCTTGTGATAGAGGAGCCCGGACGTTGGTATCAGAGTACC 1045
 151 GTGACCGCTATGGGATGTGATGATGATGATGATGATGATGATGATGATGATGAT 210
 1046 A---CTACATCCCGAGACTCAGAAAGTACAGAAAGACAACTAAGAAATGATGACA 1102
 211 AATTCTTCAACAAGCTTTCTGTGATCAGCAACATGAGGATGAGAAAGATGGCGTA 270
 1103 GACACAAAGACATCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1162
 271 GGGAGAGAGACACCCCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 330
 1163 ATCCAGAGCGGAGGATTTCTTCAAGCCCTTTATGAGATCCCAATGAATGATGAGA 1222
 331 TAGAATGTGCTGCTACTACTGAAAGATGATGATGATGATGATGATGATGATGATGAT 390
 1223 CAGCGCCGCTTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1282
 391 CCGGCGCTGTGACCTTTGTGAGCAACTTACTATGACA-----CAGACAAAG 438
 1283 CGCTAAGTACTTGCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1342
 439 GGGCTCCGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498

QY 1343 GAAGATAGAGAGAGAGCTTCAAGCTCTGGAAGAAAGCATAGAGAGCTTATGCAAGC 1402
 DB 499 GGACCTGAGAGTGTGATTCAGCTGCGAGCTGGCCACCAAGTTGAGAACTGTAAAGAACT 558
 QY 1403 ACAGAAAGCCATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1462
 DB 559 ATCAGAGCCATTTTATTCAGAGCGAGTATGAGCGAGAAACGATTCAGAGGTTTACACAG 618
 QY 1463 ATCCAGCTTCAATGTTCTCCGAGAGTACCAAGAGAGCTCTTGAAGAG-ACGATCAG 1521
 DB 619 ATCCAGCTTCAATGTTCTCCGAGAGTACCAAGAGAGCTCTTGAAGAG-ACGATCAG 678
 QY 1522 CTCCTTTGAAAGAG-----ACTCATCATGATGAGAAACAGCTGTGGGCTTTGCAAGAT 1576
 DB 679 GTCTGATCTCAAAAGCGAGAAATACGTGTGTGAGAGAGCTCATTTGGAATTTGCGGAT 738
 QY 1577 TTAAGACTCTCAGAAATGTGAGAGAGCCATTCACCAAGAGGATGTTTTCAC 1631
 DB 739 TATGATGAGAGCTGACCGAGAGAGTGTGGGAAATTAAGAGGAGATCTC 793

RESULT 14 914 bp mRNA linear EST 05-FEB-2003
 CB203472
 LOCUS CB203472
 DEFINITION AGENCOURT_11297304 NIH_MGC_135 Mus musculus cDNA clone
 IMAGE:30141354 5', mRNA sequence.
 CB203472
 ACCESSION CB203472.1 GI:28239894
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 914)
 NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Cloning Strategy: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 http://image.llnl.gov
 Plate: NDA0048 row: n column: 19
 High quality sequence stop: 580.
 Location/Qualifiers
 1..914
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30141354"
 /lab_host="DHI0B (phage-resistant)"
 /clone_1ib="NIH MGC 135"
 /note="Vector: PCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Normalized full-length enriched library from pooled mouse
 embryonic limb, maxilla and mandible, day 12.5, 13.5,
 14.5, and 15.5 (size selected for the 0.5-1 kb fragments)
 Cloned directionally, priming method: Oligo-dT. cDNA
 enrichment: >1k bp. Average insert size 1.6k bp.
 Normalization (Cot value): 7.5 kb. Priming sequence:
 5'-GACTAGTCTAGATGCGGCGGCGGCC(T)3' Tissue contributed
 by, David Rowe. Library constructed by Resgen, Invitrogen
 Corp."

ORIGIN
 Query Match 7.7%; Score 129.8; DB 6; Length 914;
 Best Local Similarity 50.4%; Pred. No. 1e-26;
 Matches 410; Conservative 0; Mismatches 382; Indels 21; Gaps 3;

QY 871 GGGCAGGGACCTTTTATCATGATGATTAAGAACTTCAACCTTCTGAAAGTGAATCAAC 930
 Db 3 GGGAAAGGTTTGATGATGGCCGCTGCTGTAAGATTTTCAACTGCTCGTTGGCTGGG 62
 QY 931 GCGAATTTTTCAGAGCTCTCACTATCTTCAAGTAAAGATGAGTCTGATCTTCCGAC 990
 Db 63 GCAAAATCTTTTCGTACCGACCACTATCCCTACTCAGAGAGATCTTCACTGTGTAC 122
 QY 991 AGACTCGGAATCTTGTGTAGACGAAGCCCGACCTTGG---TATCAAGATTAACAC 1047
 Db 123 CATATCGGATTTGTGTATCATATGATGTCTCCGGTGGGATTTGTGTACTTCAAGT 182
 QY 1048 TACAATCCCGAGACTCAGAAAGATAGCAGAAGCAACATTAAGAAATGATTCACAGAC 1107
 Db 183 TTTGGCAACGATCACTTGGGACCACTTAAGAGTATGAGAGACTGTTGGCGGAC 242
 QY 1108 AAGAACCATCCAGTGTATCATATGTAGATGTGGGCAAGCAACAGAGTCAACATCCA 1167
 Db 243 AAAAAATCACCTGCGGTGTGTATGTGTGTGTGGCCAATGACCTTCTCTCTGAAA 302
 QY 1168 GACGGGAGGGTTTCTTCAAGCCCTTTTATGAGCTGCCAATGAATGATGAAACAGC 1227
 Db 303 CCGGCGCATATTAATCTTAAAGCGTGAATCACCAACCAAGCCCTGACCTCAACCGT 362
 QY 1228 CCGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
 Db 363 CCGGTGACCTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 410
 QY 1288 AAGTACTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347
 Db 411 CCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 470
 QY 1348 ATAGAAGAGAGCTTCAAGCTTGGGAAAGATAGAAAGATCTATGATGATGATGAT 1407
 Db 471 TTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 530
 QY 1408 AAGCCATCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1467
 Db 531 AAGCGATTTATCAGAGGAGATGATGATGATGATGATGATGATGATGATGATGAT 590
 QY 1468 CCTCAATTTTCTCGAAGAGTACCAAGAGAGCTGTTGAAA-----AGACATCAG 1521
 Db 591 CTTCCATGTTTCAAGAGAGTACCAAGAGAGCTGTTGAAAATTAACATTCAGTTCTG 650
 QY 1522 CTCCTTTTGAAGAAAGATACATCAATCGGACACACGCTGTGGCCCTTGGAGATTTAAG 1581
 Db 651 GATCGAAGAAAGTAAAGATAGTGGTGGAGAGCTCATCTGAAATTTGGCGACTTCAAG 710
 QY 1582 ACTCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1641
 Db 711 AGGAACAGTCAACCGTGTAGATTAATCGGGAACAAAGAGGATCTTCACTGCGCAGAG 770
 QY 1642 CAACCAAACTGCTGCTCATGATGATGATGATGATGATGATGATGATGATGAT 1674
 Db 771 CAGCCCAAAACTTGGCTTTTATTTTGGCAGAG 803
 RESULT 15
 CO394571 862 bp mRNA linear EST 01-JUL-2004
 LOCUS CO394571
 DEFINITION AGENCOURT 27822917 NIH MGC 252 Rattus norvegicus cDNA clone
 IMAGE:7307606 5', mRNA sequence.
 ACCESSION CO394571
 VERSION CO394571.1 GI:49576487
 KEYWORDS EST.
 ORGANISM Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 862)
 AUTHORS NIH-MGC <http://mgi.mgi.nhl.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gsapbs-remail.nih.gov
 Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
 College of Wisconsin
 cDNA Library Preparation: Express Genomics
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1AM15349 row: e column: 12
 High quality sequence scop: 685.
 Location/Qualifiers
 1..862
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="IMAGE:7307606"
 /tissue_type="Ovary, strain - Norway Line3. Ages weeks.
 Tissue was snap-frozen adn transferred in -70C. RNase free
 for the entire procedure"
 /lab_host="DH10B TonA"
 /clone_lib="NIH_MGC_252"
 /note="Organ: ovary; Vector: pDONR 201; Site 1: NotI;
 Site 2: NotI; RNA obtained from female ovaries animals at
 8 wk old. Tissues were snap-frozen and kept at -80C for
 two days before RNA extraction and purification
 (TRI-reagent method). cDNA was primed using oligo-dT
 primer: 5'-pGATGTTCTTGAATGCGAGCGGCGGCC(T)25-3' and
 cloned into the EcoRV/NotI sites of pExpress-1.
 Size-selection >1.25kb resulted in an average insert size
 of 1.7kb. This primary library is not normalized
 (normalized library is NIH MGC 252) and was constructed by
 Express Genomics (Frederick, MD). Note: this is a NIH_MGC
 library"

ORIGIN
 Query Match 7.6%; Score 128.2; DB 7; Length 862;
 Best Local Similarity 51.5%; Pred. No. 3e-26;
 Matches 354; Conservative 0; Mismatches 318; Indels 15; Gaps 2;
 QY 855 GGAATTTCCCGTTCTGGGGGACGGGACCTTTATTCATGATTAAGAACTTAACCT 914
 Db 7 GAATTTCCCGGATGAGGAGAGAGGCTTCGACTGCTGTGATTAAGATTTCAACCT 66
 QY 915 TGTGAAGTGAATCAACGGGAATCTTTTCAAGACCTTCATATCTTCAAGTGAAGATG 974
 Db 67 CTTCCGTTGGCTCGGGGCAAAATTCCTTTCTGACACGACATATCTTCTGAGAGAGAT 126
 QY 975 GCTGATTTTCCGACAGACTCGAATCTTGTATGATGACGAAGCCCGCAAGTTGTAT 1034
 Db 127 ACTTCAGCTCTGTAGCCGATATGGAATGTGTATGATGATGATGATGATGATGATGAT 186
 QY 1035 CACAAGTACCA---CTACATTCGGAGACTCAGAAATAGCAGAAAGCAACATTAAG 1091
 Db 187 CGTGTGCCCAAGATTGTCGAACGATCTCTTCGACCAACCACTAAGGATGATGACGA 246
 QY 1092 AATGATGACAGACCAAGAACATCCAGTGTATCTTGTAGATGTGGAGTGGGAACGAC 1151
 Db 247 GCTGTGGCGAGGAGCAAAATTCACCTCGGCTGTATGTGTGTGTGTGTGTGTGTGTGT 306
 QY 1152 AAGTCCAAACCATCCAGACGGAGGGTTTCTTAAGCCCTTTATGAGACTGCCAATGA 1211
 Db 307 TGTCTCTTCTGTGAACCTCGCGATATTAATTCGAACGCTGATTCGCCCAACAAAGC 366
 QY 1212 AATGATGAACAGCCCGCTTGTATGATGATGATGATGATGATGATGATGATGATGAT 1271
 Db 367 CTTGAGCCCAACCGCTCTGTGACCTTTGTAGCAATA-----CCAGATATGA 414

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 17:36:30 ; Search time 330 Seconds
(without alignments)
8374.762 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689
Sequence: 1 atggttaagaccgcaacgaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	4	US-09-270-957-14 Sequence 14, Appl
2	172.4	10.2	1956	4	US-09-715-858-1 Sequence 1, Appl
3	172.4	10.2	2191	3	US-09-039-555B-12 Sequence 12, Appl
4	159.6	9.4	1947	4	US-09-715-858-3 Sequence 3, Appl
5	141.4	8.4	1887	3	US-09-149-727-7 Sequence 7, Appl
6	130.4	7.7	1888	4	US-09-270-957-27 Sequence 27, Appl
7	112	6.6	14683	2	US-08-819-866-1 Sequence 1, Appl
8	112	6.6	14683	3	US-09-023-715-1 Sequence 1, Appl
9	112	6.6	14683	3	US-09-343-485A-1 Sequence 1, Appl
10	110.8	6.6	3035	3	US-08-723-485A-1 Sequence 18, Appl
11	110.8	6.6	3169	3	US-08-630-820-5 Sequence 5, Appl
12	110.8	6.6	3169	4	US-09-819-866-2 Sequence 9, Appl
13	110.8	6.6	3824	2	US-08-723-624-19 Sequence 19, Appl
14	110.8	6.6	4652	4	US-09-893-525-36 Sequence 36, Appl
15	110.8	6.6	5390	4	US-09-893-525-41 Sequence 41, Appl
16	110.8	6.6	5418	4	US-09-893-525-38 Sequence 38, Appl
17	110.8	6.6	11978	4	US-09-792-568-8 Sequence 8, Appl
18	110.8	6.6	12438	4	US-09-792-568-9 Sequence 9, Appl
19	110.8	6.6	18986	2	US-08-819-866-2 Sequence 2, Appl
20	110.8	6.6	18986	3	US-09-023-715-2 Sequence 2, Appl
21	110.8	6.6	18986	3	US-09-343-485A-2 Sequence 2, Appl
22	110.8	6.6	32798	4	US-09-604-954B-1 Sequence 1, Appl
23	109	6.5	1806	4	US-09-270-957-7 Sequence 7, Appl
24	109	6.5	1854	4	US-09-149-727-3 Sequence 3, Appl
25	109	6.5	2100	4	US-09-270-957-1 Sequence 1, Appl
26	109	6.5	6029	3	US-09-149-727-1 Sequence 1, Appl
27	106.8	6.3	1503	4	US-09-949-016-5826 Sequence 5826, Ap

28	104.4	6.2	8012	3	US-09-182-117-1 Sequence 1, Appl
29	104.4	6.2	8012	4	US-09-434-039A-1 Sequence 1, Appl
30	104.4	6.2	8418	3	US-09-182-117-5 Sequence 5, Appl
31	104.4	6.2	8418	4	US-09-434-039A-5 Sequence 5, Appl
32	104.4	6.2	8798	3	US-09-182-117-4 Sequence 4, Appl
33	104.4	6.2	8798	4	US-09-434-039A-4 Sequence 4, Appl
34	98	5.8	2141	4	US-09-445-283C-27 Sequence 27, Appl
35	98	5.8	2633	3	US-08-452-267-2 Sequence 2, Appl
36	98	5.8	2633	1	US-09-123-644-2 Sequence 2, Appl
37	98	5.8	4544	4	US-09-488-270A-1 Sequence 1, Appl
38	98	5.8	4947	3	US-09-118-276-21 Sequence 21, Appl
39	98	5.8	5534	1	US-08-452-267-3 Sequence 3, Appl
40	98	5.8	5534	3	US-09-123-644-3 Sequence 3, Appl
41	98	5.8	5560	3	US-08-817-188-5 Sequence 5, Appl
42	98	5.8	5642	1	US-08-318-772A-2 Sequence 2, Appl
43	98	5.8	5897	3	US-09-097-319A-26 Sequence 26, Appl
44	98	5.8	5897	4	US-09-643-971-26 Sequence 26, Appl
45	98	5.8	6898	3	US-09-097-319A-27 Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-270-957-14
; Sequence 14, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270.957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PaedSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1689)
; OTHER INFORMATION: n = A,T,C or G
US-09-270-957-14

Query Match	99.9%	Score 1688	DB 4	Length 1689
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1689	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	1	ATGTTAAGACCGCAACGAAGAGATTATTTCTTGAATGAGATTGGAT	60	
DB	1	ATGTTAAGACCGCAACGAAGAGATTATTTCTTGAATGAGATTGGAT	60	
QY	61	CTTGAAGTACCGAAGAGACCAATCGCGTCTTGAAGTGAAGAGAGTAC	120	
DB	61	CTTGAAGTACCGAAGAGACCAATCGCGTCTTGAAGTGAAGAGAGTAC	120	
QY	121	CAGGATCTGTGCTACGAGAGAGACCTTCACTTCAAAACCACTTACGTTCCGAG	180	
DB	121	CAGGATCTGTGCTACGAGAGAGACCTTCACTTCAAAACCACTTACGTTCCGAG	180	
QY	181	NAACCTTCAAAAACATCATGACTTCTTCTGCGGTGACACGAGCTGCGAGTTC	240	
DB	181	NAACCTTCAAAAACATCATGACTTCTTCTGCGGTGACACGAGCTGCGAGTTC	240	
QY	241	TTCCCTCAACGAGAGAAAGTGGAGAGATCATATTGCTTCCCTTGAATAGAT	300	
DB	241	TTCCCTCAACGAGAGAAAGTGGAGAGATCATATTGCTTCCCTTGAATAGAT	300	
QY	301	GTCAGCGGAGAAAGTGAATCCGAGAGAGCAACTCAGGCTGTTGAGAAAGATTG	360	
DB	301	GTCAGCGGAGAAAGTGAATCCGAGAGAGCAACTCAGGCTGTTGAGAAAGATTG	360	

```

QY 361 AAAGTGGAGATTTTCCCTCGAAGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT 420
Db 361 AAAGTGGAGATTTTCCCTCGAAGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT 420
QY 421 GGAAGTTTCCACCTGCAAACTTCCACTTCCCTCAGTGAATCATAGAGCTGT 480
Db 421 GGAAGTTTCCACCTGCAAACTTCCACTTCCCTCAGTGAATCATAGAGCTGT 480
QY 481 CTGATAGATTCAACAGACCGGAGATTACTGACATCTGGGTGACACGAGTAGTCT 540
Db 481 CTGATAGATTCAACAGACCGGAGATTACTGACATCTGGGTGACACGAGTAGTCT 540
QY 541 GAAACCGGAGAAATTTGGAAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGACGGTG 600
Db 541 GAAACCGGAGAAATTTGGAAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGACGGTG 600
QY 601 GGAACAGAGATGACGATCAAACTTGGAGAGAAAGAAAAAGTTAGAACATCCAACAG 660
Db 601 GGAACAGAGATGACGATCAAACTTGGAGAGAAAGAAAAAGTTAGAACATCCAACAG 660
QY 661 TTGTCGAAAGGAGATTCAATCTCGAAAAAGCCAGGTTCTGAGCCTCGAAGATCCATAT 720
Db 661 TTGTCGAAAGGAGATTCAATCTCGAAAAAGCCAGGTTCTGAGCCTCGAAGATCCATAT 720
QY 721 CTTTATCCTCTCAAGGTGGAATTTGAAAAAGACGATACATCTGACATCGGAATTCGA 780
Db 721 CTTTATCCTCTCAAGGTGGAATTTGAAAAAGACGATACATCTGACATCGGAATTCGA 780
QY 781 ACAGTACGTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 840
Db 781 ACAGTACGTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 840
QY 841 TTTGGAACGACGAGAAATTTCCCGTTCTGGGGCAGGGCACTTTTATCATTTGATGATA 900
Db 841 TTTGGAACGACGAGAAATTTCCCGTTCTGGGGCAGGGCACTTTTATCATTTGATGATA 900
QY 901 AAAAGATTTCAACTTTGAAAGTGAATCAACGGAAATTTCTTGAAGACCTCAGATTCCT 960
Db 901 AAAAGATTTCAACTTTGAAAGTGAATCAACGGAAATTTCTTGAAGACCTCAGATTCCT 960
QY 961 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGCTCGGAATCTTGTGATAGACGAGCC 1020
Db 961 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGCTCGGAATCTTGTGATAGACGAGCC 1020
QY 1021 CCGCAGTTGGTATCAAAAGTACCACTCAATCCGAGACTTCAGAAATAGCAGAAAGAC 1080
Db 1021 CCGCAGTTGGTATCAAAAGTACCACTCAATCCGAGACTTCAGAAATAGCAGAAAGAC 1080
QY 1081 AACATGAAGAAATGATCGACAGACCAAGAACCAATCCAGTGTATCATTTGAGAGTGTG 1140
Db 1081 AACATGAAGAAATGATCGACAGACCAAGAACCAATCCAGTGTATCATTTGAGAGTGTG 1140
QY 1141 GCGAAAGAAACAGAGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAAACCTTTATGAG 1200
Db 1141 GCGAAAGAAACAGAGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAAACCTTTATGAG 1200
QY 1201 ACTGCCAATGAATGATGAGAACACGCCCCGTTGTATGATGATGATGATGATGATGATG 1260
Db 1201 ACTGCCAATGAATGATGAGAACACGCCCCGTTGTATGATGATGATGATGATGATGATG 1260
QY 1261 GACGAGAGAAACAGAGACGTTGGGCTGGAAGTATTTGACATCTGTGTGAAACAGGTAC 1320
Db 1261 GACGAGAGAAACAGAGACGTTGGGCTGGAAGTATTTGACATCTGTGTGAAACAGGTAC 1320
QY 1321 TACGGCTGGTATCATCTATCAGGAAAGATGAGAAAGGACTTCAAGCTCTGAAAAAGAC 1380
Db 1321 TACGGCTGGTATCATCTATCAGGAAAGATGAGAAAGGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGAC 1440
Db 1381 ATGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGAC 1440

```

```

QY 1441 GCGATAGCTGACATCCATAGATTCACCTCAATGTTCTCCGAAGATACCAAGACAG 1500
Db 1441 GCGATAGCTGACATCCATAGATTCACCTCAATGTTCTCCGAAGATACCAAGACAG 1500
QY 1501 CTCGTTGAAAAAGACATCAGGCTCTCTTTTGAAAAAAGATACATCATCGAAACACG 1560
Db 1501 CTCGTTGAAAAAGACATCAGGCTCTCTTTTGAAAAAAGATACATCATCGAAACACG 1560
QY 1561 TGGGCTTTGCAAGATTTTAAAGATCTCTCAGAAATGAGAAAGACCATTTCTCAACAG 1620
Db 1561 TGGGCTTTGCAAGATTTTAAAGATCTCTCAGAAATGAGAAAGACCATTTCTCAACAG 1620
QY 1621 GGTGTTTCAACAAGACAGACCAACCACTCGTTCATGATGAGAAAGACGTTG 1680
Db 1621 GGTGTTTCAACAAGACAGACCAACCACTCGTTCATGATGAGAAAGACGTTG 1680
QY 1681 AGTGAAGTT 1689
Db 1681 AGTGAAGTT 1689

RESULT 2
US-09-715-858-1
; Sequence 1, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYSSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715, 858
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
US-09-715-858-1

Query Match 10.2%; Score 172.4; DB 4; Length 1956;
Best Local Similarity 52.1%; Pred. No. 5.5e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

Qy	1094	TGATCGACAGACAGAAAGACCATCCAGATGTGATCATGTGAGAGTGTGGCGAAACGAG	1153
Db	1298	TGTGTGCGTAGGAGCAAGAACCAACCCCGGGTCTGTGATGTGGTCTGTGTGCGCAACGAGCTG	1357
Qy	1154	AGTCCAAACCATCCAGACGCGGAGGGTTCCTTCAAGCCCTTTATGAGCTGCCAATGAA	1213
Db	1358	CGTCCCACTAGAAATCTGTGGCTACTCTTGAAGATGTATGATGGCTCACACCAATTCCT	1417
Qy	1214	TGATTCGAACACGCCCTTGTCTATGTGTGACATGATGAGACGACACGACAGAAACAA	1273
Db	1418	TGAGACCCCTCCCGGCTGTGTGACTTTTGTGACCAACTCTTAATTATGACG-----	1465
Qy	1274	GAGACGTGGCCCTGAAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTACA	1333
Db	1466	CAGACAGAGGGGGCTCCGTATGTGATGTGATCTGTGTGAACAGCTACTACTCTTGTGTATC	1525
Qy	1334	TCTATCAGGGAAGATAGAAAGAGACTTCAGCTCTTGAAAAAAGACATAGAAGCTCT	1393
Db	1526	ACGACTACGGGCACTGTGAGTTGATTCAGCTGCAGCTGGCCACCACTTTGAGAACTGTGT	1585
Qy	1394	ATGCAAGGCACAGAAAGCCCATCTTTGTGCACAGAAATTCGGTCCGACGCGCATTACTGGCA	1453
Db	1586	ATAAGAAATATCAGAAAGCCATTTATTCAGACGAGTATGAGACGAAACGATTCAGAGGT	1645
Qy	1454	TCCACTACGATCCACCTCAATGTCTTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1646	TTCAACAGGATTCACCTCTGTGATGTTCATCTGAAGAGTACCAAAAAAGTCTGTAAAGCAGT	1705
Qy	1513	ACGATCAGGCTCCTTTTGAAAAA-----AGACTATCATCATCGAAACACGCTGTGGGCT	1567
Db	1706	ACCATCTGTGGTCTGGATCAAAAAACGCAAAAAATATGTGTGGAGAGCTCATTTGGAATT	1765
Qy	1568	TTGCAGATTTTAAAGCTCCTCAGATGTGAGAAAGCCATTCACACCAAGGCTGTTT	1627
Db	1766	TTGCCGATTTATGATGCTGAACAGTCACCGACGAGAGTGTCTGGGAATATAAAAAGGAGTCT	1825
Qy	1628	TCACAAAGACAGACAAACCCAAACTCGTTGCTCATGTACTGAGA	1671
Db	1826	TCACTCGGCAAGACAAACCAAAAGTGACGCTTCCTTTTCGA	1869
RESULT 3			
US-09-039-555B-12			
Sequence 12, Application US/09039555B			
Patent No. 6033856			
GENERAL INFORMATION:			
APPLICANT: Koerner, Kathrin			
APPLICANT: Mueller, Rolf			
APPLICANT: Sadlasek, Hans-Harald			
TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS			
TITLE OF INVENTION: PREPARATION AND USE			
NUMBER OF SEQUENCES: 19			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Foley & Lardner			
STREET: 3000 K Street, N.W., Suite 500			
CITY: Washington			
STATE: D.C.			
COUNTRY: USA			
ZIP: 20007-5109			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patentin Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/039, 555B			
FILING DATE: 16-MAR-1998			
CLASSIFICATION: 514			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: DE 19710643.9			
FILING DATE: 14-MAR-1997			
ATTORNEY/AGENT INFORMATION:			

:	NAME:	Bent, Stephen A.
:	REGISTRATION NUMBER:	29,768
:	REFERENCE/DOCKET NUMBER:	016779/0131
:	TELECOMMUNICATION INFORMATION:	
:	TELEPHONE:	(202) 672-5300
:	TELEFAX:	(202) 672-5399
:	TELEX:	904136
:	INFORMATION FOR SEQ ID NO:	12:
:	SEQUENCE CHARACTERISTICS:	
:	LENGTH:	2191 base pairs
:	TYPE:	nucleic acid
:	STRANDEDNESS:	single
:	TOPOLOGY:	linear
:	MOLECULE TYPE:	DNA (genomic)
:	US-09-039-555B-12	
Query Match	10.2%;	Score 172.4; DB 3; Length 2191;
Best Local Similarity	52.1%;	Pred. No. 5.9e-43;
Matches	492; Conservative	0; Mismatches 431; Indels 21; Gaps 4;
OY	737	TGGAATTGAAAAAGACGAGTCACTGTGGACATCGAAATCAGAACAATCAAGCTGAGGACG 796
Db	964	TGGGGCCCTGTGTCTGACTTCTACACACATCCCTGTGGGAGATCCGCATGTGGCTGTCA 1023
OY	797	AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAAGGGCTTTGGAAAGCAGAGG 856
Db	1024	AGACCACAGTTCCTCATCAATGGGAAACCTTTCTATTTCACAGGTGTCAACAGCATBAAG 1083
OY	857	AATTCGCCGTTCTTGGGCGAGGCGACCTTTATCCATTGATGATATAAGACTTCAACCTTC 916
Db	1084	ATGGGACATTCGAGGGAAAGGCTTGCATGCGCGCTGTGTGTGAAGACTTCAACCTGCG 1143
OY	917	TGAAGTAGATCAACGCGAATTCTTTACAGACCTTCACTATCTTACAGTGAAGATGGC 976
Db	1144	TTCCCTGGCTTGTGGCCAACGCTTCCGTAACAGCCACTACCCCTATGAGAGAAATGA 1203
OY	977	TGGAATCTTGCCGACAGACTGGAACTCTTGATGATAGCGAAGCCCCGACGTTGGATCA 1036
Db	1204	TGCAGATGTGTGACCGCTATAGGATGTGTGATCATCATGATGTGTCCGCGTGGGCCCTGG 1263
OY	1037	CNAGGTACCA---CTAACATCCGAGACTCGAAGATAGGAGAAACAACTAAGAAAGAA 1093
Db	1264	CGCTGCCGCAATTTTCAACAACGTTCTCTGCATCACACATGCAAGTATGGAABAG 1323
OY	1094	TGATCGACAGACACAGAACCATCCAGTGTGATCATGTGGAGTGTGGCGAACGACACAG 1153
Db	1324	TGTGTCGATAGGAGCAAGAACCAACCCCGCGGTGTGATGTGGTCTGTGGCCAAAGAGCTG 1383
OY	1154	AGTCCAAACATCCAGACGCGGAGGGTTTTCTTCAAAGCCCTTTATGAACTGCCATGAAA 1213
Db	1384	CGTCCCACTTGAATCTGTCTGGCTACACTTGAAGATGATGTGCTCACACCAAAATCTT 1443
OY	1214	TGATTCGAACAACGCCCGGTGTGTCATGTGGACATGATGAGACGACACGAGAGAAACAA 1273
Db	1444	TGGACCCCTCCCGGCTGTGACCTTTTGTAGCACTTAACATATGACG----- 1491
OY	1274	GAGACGTGGCGCTGAAGTACTTTCGACATCTGTGTGTGAACAGGTACTACCGCTGTACA 1333
Db	1492	CAGACAAAGGGGCGTCGGTATGTGATGTGATCTTTTGAACAGTACTTACTCTTGGTATC 1551
OY	1334	TCATTCAGAGGAGATAGAGAGAACTTCAAGCTCTTGAAAAAGACATAGAGAGCTCT 1393
Db	1552	ACGACTACGGGCACTGTGAATGTATTCAGCTGCAGCTGGGCCACCACTTTTGAAGAACTGTG 1611
OY	1394	ATGCAAGGCAACAGAAAGCCCATCTTTTTCACAGAAATCGGTTGGGAGCGCATATGCTGGCA 1453
Db	1612	ATTAAGAGTATCAGAGGCCATTATTCAGAGCGGTATGAGACAGAAACGATTTCAAGGT 1671
OY	1454	TCCACTACGATCCACCTCAATGTTCTTCGAAAGTACCAAGCAGAGCTGTTGAAAAAG- 1512
Db	1672	TTTCAACAGGATCCACCTCTATGTTCATGAAGATACAGAAAAGCTGTGTGAGACAGT 1731
OY	1513	ACGATCAGGCTCTTTTGA AAAA----AGACTATCATCATCGAACAACAGTGTGGGCTT 1567

```

Db      1732 ACCATCTGGGTCTGATCAAAAACGAGAAAATATGTGTTGGAGAGCTCATTTGAAAT 1791
Qy      1568 TTGCAGATTTTAAAGTCTCTGAGATGTGAAGAACCCATTCTCAACCAAGGTTT 1627
Db      1792 TTGCCATTTCTAGCTGAACAGTCAACGACGAGTGTCTGGGATATAAAAGGGATCT 1851
Qy      1628 TCACAAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
Db      1852 TCACCTGGCAGACACAAACCAAGTGCAGCTTCTTTGGCA 1895

```

RESULT 4

```

US-09-715-858-3
; Sequence 3, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VARIANTS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYSSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715,858
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1947)
US-09-715-858-3

```

Query Match Beel Local Similarity 9.4%; Score 159.6; DB 4; Length 1947;

Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

```

Qy      739 GAACCTGAAAAAGACGATACACTCTGACATCGAATCAGAACGATCAGTGGACGAG 798
Db      928 GAGTCTGATGACTGACTACTACACCTTCTTCTCGGGATTCGAACAGTGTCTGTCAAAAG 987
Qy      799 AAGAGCTCTATCTGAACGGGAACTGTCTTTTGAAGGGCTTTGAAAGACGAGAA 858
Db      988 AACAACTTCTCATTAACGGGAAGCCCTTCTATTTTCAAGGGGTCAATTAAGACGAGAT 1047
Qy      859 TTCCCGCTCTGGGGGACGACCTTTTATCCATTGATGATTAAGACTTCAACCTTCTG 918
Db      1048 TCAGATATTCGAGGGAAGAGCTTCCAGCTGGCCGCTCTGTAAAGATTTCACCTGCTC 1107
Qy      919 AAGTGATCAACGCAATTTCTTTCAGACCTCTCACTATCTTACAGTGAAGTGGCTG 978
Db      1108 CGTTGGCTCGGGCAAAATTCCTTTCGTAACAGCACTATCCCTACTCAGAGAGGTACTT 1167
Qy      979 GATCTTGGCAGACGACTCGGAATCTTGTGATGAGAGAGCCCGCAGCTTGG--TATC 1035
Db      1168 GAGCTCTGTGACCGAATCGGAGTTGTGATCGAATGTGCCGTGTGGGCAATTTG 1227
Qy      1036 ACAAGGTACCACTACAAATCCGAGACTCAGAAAGATGAGAGAACAACTAAGAAATG 1095
Db      1228 CTACCTCAGAGTTTGGCAAGAGTCACTTCGACACACTAGAGGTATGAGAGAGCTG 1287
Qy      1096 ATCGACAGACACAAGAACCATCCAGTGTGATGATGAGTGTGGCGAACAACGAG 1155
Db      1288 GTTCGCGGGAACAATAATCACTCGGTGTGATGTGTGTGTGGCAATGAGCTTCC 1347
Qy      1156 TCCAAACATTCAGACCGGAGGTTTCTTCAAGCCCTTTAAGACTCCCAATGAATG 1215
Db      1348 TCTGCTTGAAACCCGCGCATATTAATTAAAGACGCTGATCAACCAACCAAGCCCTG 1407

```

```

Qy      1216 GATCGAACCGCCCGTTGTCTATGTTGAGCATGATGACGACCGACGAGAACAG 1275
Db      1408 GACCTACCCCTCCCGTCACTTTGTGACA-----ACGCCAAATATATATGCA 1455
Qy      1276 GACGTGGCGCTGAAGTACTTGACATCTGTGTGTGAACAGATTAAGGCTGTATC 1335
Db      1456 GACCTGGGGGCCCGGTATCGTGAATGTATCTGTGTAACAGCTACTTTTCTTGGTATCAT 1515
Qy      1336 TATCGAGGAAGATTAAGAAAGAGCTTCAAGCTCTGAAAAAAGATGAAGAGCTTAT 1395
Db      1516 GACTATGGGCATTTTGAAGTATTCAGCCACAGCTGAATACCACTTTGAGAATCTGAT 1575
Qy      1396 GCAAGGACAGAAAGCCCATCTTTGTACAGAAATTCGTTGGGAGCGCATAGCTGCATC 1455
Db      1576 AAGAGCATTCAGAAAGCCATTTATCCAGGCGAGTATGAGACGACGCAATCCAGGGATC 1635
Qy      1456 CACTAAGTCCACCTCAAAATGTTCTCCGAGAGTACCAAGACAGAGCTGTGAAAAAG 1515
Db      1636 CACGAGACCCGCGCTCGCATGTTGAGTGAAGATCCAGAAAGGCTGTTCTGGAATTTAC 1695
Qy      1516 ATCAGGCTCTTTTGAAA-----AAGACTATCATTCGGAACACAGCTGTGGCTTT 1569
Db      1696 CATTCAGTTCTGATTCAGAAAAAGTAAAGAAATACGTGTGGAGAGCTCATCTGGAATTT 1755
Qy      1570 GCAGATTTTAAAGCTCTCAGAAATGTGAAGAACCCATTTCTCAACCAAGGTTGTTTC 1629
Db      1756 GCCACTTCATGACGAAACCACTACCGCTGTGAGAGTATCGAAGAACAAAGGAGATCTTC 1815
Qy      1630 ACAAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGA 1671
Db      1816 ACTGCGCAGACAGACGCCAAAACTTCGCGCTTTATTTTGGCA 1857

```

RESULT 5

```

US-09-149-727-7
; Sequence 7, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Killen, Andrew
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1887
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-149-727-7

```

Query Match Beel Local Similarity 8.4%; Score 141.4; DB 3; Length 1887;

Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

```

Qy      120 CCAGATCTGTCTTACGAAGAAGACCTTCACTCAAAAACCACTTCTACGTTCCGA 179
Db      208 CAAAGAAATCCGCAACATATCGATATGTCTGTGACGAAGTAGATTACAGGTCCCGC 267
Qy      180 GNACTTTCAAAAACATCACTCACTTTACTTTGTCGCTGGAACACGACCTGCAAGT 239
Db      268 CTATCTGAAGATCAGCTATCTGCTCCGCTCGGCTCTCAACTCAAAAGCAATTTGT 327
Qy      240 CTCTCTCAACGAGAAAGTGGAGAAATCACTAATTGAATACCTTCCCTTCAAGTGA 299
Db      328 CTATGTCAATGATGAGTGTGTGTGTGAGCAACAAGGGCGAATTCCTGCAATTCGAAGCGGA 387
Qy      300 TGTGACGGGAAAGTGAATCCGAGAGAAAGAACTCAGGGTGTGTTGTGAAACAGATT 359

```

```
Db      388 AATCAACAACCTGCGTCGTGATGGCATGATGCGGCTCACGTCGCGGTGGACAACATCT 447
Qy      360 GAAAGTGGAGAGATTTTCTCGAAGTTTCAGACAGCGGCACCTCACACCGTGGATTTT 419
Db      448 CGACGATACACCCCTCCCGGTGGGCTGTACAGCGGCGCACAGAGAGGCGCTCGGAA 507
Qy      420 TGGAAATTTTCCACCTGCAACCTTGACCTTTCCCTACGCGTGAATGATAAGCGCTGT 479
Db      508 AGTCATTCCTTAACAGCCGAACTTGACCTTTCAACTATGACAGGCTCAGACCGTCCGT 567
Qy      480 TCTGATAGATTCAACAGACGACGAGATCTGACATCTGGGTGACAGCAGATGATGTC 539
Db      568 GAAATCTACACGACCCCGTTTACGTACGTGAGGACATCTGGTTGTGACCGACTTCA 627
Qy      540 TGAACCGGAGAAAGAACTTGAAAGAGTGAAGTGAAGTCAAGAAAGCGGT 599
Db      628 TGGCCCAACCGGGACTGTACCTATACGGTGAAGCTTTCAAGGCAAGCCGAGACCGTGA 687
Qy      600 GGGACGAGAGATGACGATCAAACTTGAGAGAGAGAAAGATTGAAACATCCACAG 659
Db      688 AGTTCGATCGTGTGATGAGAGAGGCAAAAGTGTGCAACACGAGGCGCTGAGCGGTA 747
Qy      660 ATTCTGCAAGGGGAGTTTCATCTCGAAAACGCAAGTTCTGAGGCTTGAAGATCCATA 719
Db      748 CGTGGAGATTCGGAATGTCTCTCTGGAACCACTGAACAGTATCTTACCAATCA 807
Qy      720 TCTTTATCTCTCAAGTGAAGCTTGAAAGAGACGATCACTCTGACATGGAATCAG 779
Db      808 AGTGAACCTGTGAAGCAGCGACTGACATCATGTCTATGAAGAGCGCTGCGGCTGCG 867
Qy      780 AACGATCAAGCTGGAGAGAGAGGCTTATCTGAACGGGAAACCTGCTTTTGAAGG 839
Db      868 GACCGTGAAGTCAACGACGAGGAGTTCTTCAACAACAAACGTTTACTTCAAGG 927
Qy      840 CTTTGGAAAGCAGAGAAATTCCTGCTGGGAGGAGGACCTTTTATCAATTGATAT 899
Db      928 CTTTGGCAAAAGTAGAGACATCTTATCAACGCGCTGCTTTAAGAACAGACATGT 987
Qy      900 AAAAGCTTCAACCTTTGAAGTGAATCAACGCGAAATTTCTTTCAGACCTCTCACTATCC 959
Db      988 GATGATTTCAATATCTTCAATATGATCGGCGCCAAACAGCTTCCGAGCGCACATATCC 1047
Qy      960 TTACAGTGAAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTTGATAGAGAGC 1019
Db      1048 GTACTGTGAAGTGTGATGCTTTCGCGATCGCAGAGGCTGTGTGATGACAGAAC 1107
Qy      1020 CCGGACGTTGGTAT----- 1034
Db      1108 TCCGGCAGTTGGCGCTGCACCTCACTTCACTGGCACACAGGAGCTCGCGAAGGACGGA 1167
Qy      1035 ---CAAGAGTACCACTACATCCGAGAGCTCAGAAAGTAGAGAAACAACTAAGAG 1091
Db      1168 GCGGCTTCAGTACTTGGAGAGAAATTCGAGCTTTGAGCACATCAAGAGCTTCTCCGTGA 1227
Qy      1092 AATGATTCAGACACAAAGAACCATCCAGTGTGATCATGTGAGTGTGGAGAAAGAAC 1151
Db      1228 ACTGTGTCTGTGCAAGAACCATTCAGAGCTGTGTGATGTGAGACATCGCAAGAGGC 1287
Qy      1152 AGAGTCCAAACATCCAGAGCGGAGGTTTCTTCAAGGCGCTTTTGAAGCTGCCATGA 1211
Db      1288 GCGGACTGAGAGAGAGGCGCTGATGAGTACTTCAAGCGCTGTGTGAGCTGACCAAGGA 1347
Qy      1212 AATGATGAG---ACACGCGCGCTGTGTCAAGTGTGAGATGATGAGCGCACAGAGAG 1268
Db      1348 ACTTCAGCCCAAGAGCGCTCCGCTACGATCTGTGCTTTGTGTGAGGTACCCCGAGAGC 1407
Qy      1269 AACAGAGACCTGGCGCTGAGTACTTGCATCTGTGTGTGAACAGTACTACAGGCTG 1328
Db      1408 GGAACAAAGTCCGCAAGCTGA-----TTGACGTCAATCGGCTCAATCGCTAATACGAGT 1461
Qy      1329 GTACATCTATCAGGAGAGATGAGAGAG---GACTTCAAGCTCTGAAAAAGACATGGA 1385
```

```
Db      1462 GTACTTCATGCGGCTGATCTTCGAAGCGGCCAAAGTCCATCTCCGCGAGAAATTTACGCG 1521
Qy      1386 AGAGCTTATGCAAGGACACAGAAAGCCATCTTTTTCACAGATTCGGTGGAGCGGAT 1445
Db      1522 GTGGAACAAGCGTTTCCAGGAAAGCCGATATGATCATGTGATCGGCGGACACCGT 1581
Qy      1446 AGCTGCATTCACATCAATCCACCTCAATATGTTCTCGAAGGTACCAAGAGACCTGT 1505
Db      1582 TCGGAGCTTTTCAGCAGATTGATTCAGTATGATCCAGAGAAATTCAGTGTGATCTTA 1641
Qy      1506 TGAAGAGAGATCAAGCTCTTTTGAAGAAAGCTACATCAATCGAACACAGTGTGGC 1565
Db      1642 CAGGCGAACACAGTCGTGTTGATGATGATGAGAACTTGTGTGGTGAAGAGCTGGAA 1701
Qy      1566 CTTTGCAGATTTTAAAGCTCTCAGATGTGAAAGAACCCATTCTCAACCAAGGCTGT 1625
Db      1702 CTTCCGGAAGCTTCGAGCTCTCAGGAGGTGTGATGCGCTCCAAAGGAAACAAAGAGGCGT 1761
Qy      1626 TTTCAAGAGAGACAGAACCAACCTGTTGCTCATGTACTGAGAGACTGTGGA 1681
Db      1762 GTTCACTGTGACCGAAGCGAAGCTCGCGGCAAGCTTTTGCAGAGCGCTGGA 1817

RESULT 6
US-09-270-957-27
; Sequence 27, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270.957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-270-957-27

Query Match      7.7%; Score 130.4; DB 4; Length 1888;
Best Local Similarity 45.4%; Pred. No. 7.7e-30;
Matches 734; Conservative 0; Mismatches 822; Indels 61; Gaps 5;

Qy      120 CCAGGATCTGTGTAAGAAAGAGACCTTCACTCAAAACCACTTCTAGTTCGAA 179
Db      208 CAAGGAATTCGCAACCATATCGATATGTCTGGTACGAAGTTCACGGTCCGCG 267
Qy      180 GNAACCTTTCACAAAACACATCAGACTTTACTTTGCTGCGGTGAACAGGACTGAGGT 239
Db      268 CTATCTGAAGATCAGCGTATCGTCTCCGCTTCCGCTCTGCAACTCAAGAAATGT 327
Qy      240 CTCTCTCAACGAGAGAAAGTGGAGAGATCAATTAATCTTCCCTTCTGAAAGTGA 299
Db      328 CTATGTCAATGATGAGCTGTGTGTGAGACAAAGAGGCGAATTTCTGCAATTCGAGCGGA 387
Qy      300 TGTGACGGGAGAAAGTGAATCCGAGAGAGAAAGAACTCAGGATGTGTGTGAGAACAGATT 359
Db      388 AATCAACAACCTGCGTCGTGATGGCATGATGCGGCTCACGTCGCGGTGGACAACATCT 447
Qy      360 GAAAGTGGAGAGATTTTCTCGAAGTTTCAGACAGCGGCACCTCACACCGTGGATTTT 419
Db      448 CGACGATACACCCCTCCCGGTGGGCTGTACAGCGGCGCACAGAGAGGCGCTCGGAA 507
Qy      420 TGGAAATTTTCCACCTGCAACCTTGACCTTTCCCTACGCGTGAATGATAAGCGCTGT 479
Db      508 AGTCATTCCTTAACAGCCGAACTTGACCTTTCAACTATGACAGGCTCAGACCGTCCGT 567
Qy      480 TCTGATAGATTCAACAGACGACGAGATCTGACATCTGGGTGACAGCAGATGATGTC 539
Db      568 GAAATCTACACGACCCCGTTTACGTACGTGAGGACATCTGGTTGTGACCGACTTCA 627
```

```

QY 540 TGAACGGAGAGAACTTGAAGAGTGAAGATAGAAGTCTCGAGAGAGCGGT 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 628 TGGCCCAACCGGGACTGTACCTTAACGTGACCTTTCAAGGCAAAAGCCGAGACCGTGA 687
QY 600 GGGACAGAGAGATGACGATCAAACTTGGAGAGAGAGAGAAAAGATTAGAACATCCACAG 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 ACTGTGCGTGTGTGTAGAGAGAGCAAAAGTGGTCGGAACCGAGAGGCGTGAAGGTA 747
QY 660 ATTGTGCAAGGGAGGATTATCTTC- GAAAAGCCAGAGTTCTGGAGCCTCGAAGATCCAT 718
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 CGTGAAGATTCGGAATGTATCTCTCGGAAACCACTGAACAGATATCTTACCCAGATCA 807
QY 719 ATCTTATCTCTCAAGGTGAACTTGAAGAAAGACAGATACATCTGACATCGAATCA 778
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 AAGTGAATCTGTGAAACGACGACATGATGATCTTAAGAACCGCTTCGCGCTGC 867
QY 779 GAACGATCAGCTGGAGCGAGAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAG 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 868 GGAACGATGAGTCAACGACGCGCAAGTTCCTCATCAACAAACCGTTCTTCAAGG 927
QY 839 GCTTTGAAAGACGAGGAATCCCGCTTCTGGGCGACCGCTTTATCATTTGATGA 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 928 GCTTTGGCAAAACATGAGGACACTCTTATCAAGCGCGTGGCTTTAAGAGGAGCAATG 987
QY 899 TAAAGACTTCAACCTTGAAGTGAATCAAGCGAATCTTTCAAGACCTCTCACTATC 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 TGAATGATTTCAATATCTTCAATATGATGGGCGCAACGCTTCGGACCGACATATC 1047
QY 959 CTTAAGTGAAGATGAGTGTGATCTTGGCAGAGACTCGGAATCCTTGTGATAGAGAG 1018
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 CTAATCTGAAAGATGATGCGCTTGGCGAGATCGGAGGCTGTGTGTGATGACGAGA 1107
QY 1019 CCCCCAGCTGTAT----- 1034
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 CTCGGCAATGTGCGGACCTCAACTTGTAGTGCACCAAGGACTCGGAGAGGACAGG 1167
QY 1035 -----ACAAGTATCACTCAATCCGAGATCAGAGATAGAGAACATATAGAA 1090
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 AGCGGTGATGATCTGGAGAGATTCGACGCTTTGAGACATCAAGACGTTCTCCG 1227
QY 1091 GAATATGACAGACAGAGAACATCCAGTGTGATCATGTGTGAGTGTGGGAGAGAAC 1150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 AACTGTGTCTGTGACAGAGAACATCCAGGCTGTGTGTGAGCATCGCAACGAGG 1287
QY 1151 CAGATGCCAATCCAGACGCGGAGGCTTTTCAAGCCCTTTATGAGATGCGCAATG 1210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 CGGCACTGAGAGAGAGGCGGTAGAGTACTTCAAGCGTTGGTGAAGTGAACGAG 1347
QY 1211 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 AACTGATGCCAAGAGAGGCTCGGCTGACATGCTGTGTGTGTGTGTGTGTGTGTGTGT 1407
QY 1268 GAACAGAGAGCTGGCGCTGAAGTACTTGACATGCTGTGTGTGTGTGTGTGTGTGTGTGT 1327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 CGGACAAAGTCCCGCAACTGA-----TTGAGTGTATCCGCTCAATCGCTATTAACGAT 1461
QY 1328 GGTATCATCTAGAGAGAGATAGAGAG---GACTTAACTCTGGAAGAAAGACATAG 1384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1462 GGTACTTGCATGAGCGGTGTATCTGGAAGCGGCAAGTTCATCTCGCCAGGAATTTCAAG 1521
QY 1385 AAGAGCTATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1522 CGTGAAGAGAGGTTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1581
QY 1445 TAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1582 TTGCGGGCTTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1641
QY 1505 TTGAAAGAGAGATGAGGCTCTTTTGAAGAAAGATGATGATGATGATGATGATGATGATGAT 1564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1642 ACCAGGCAACCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1701

```

```

QY 1565 CTTTGCAGATTTTAAAGCTCTCAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1702 ACTTCGCGGACTTGGCAGACCTCTCAGAGCGGTGATGCGGTGATGCGGAGAGAGAGAGAGAG 1761
QY 1625 TTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1762 TGTTCATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818

```

RESULT 7

```

US-08-819-866-1/c
; Sequence 1, Application US/08819866
; Patent No. 5830698
; GENERAL INFORMATION:
; APPLICANT: REF. Mitchell E.
; APPLICANT: BARNETT, Richard Spence
; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
; TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
; TITLE OF INVENTION: RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,866
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-352
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-819-866-1

```

```

Query Match 6.6%; Score 112; DB 2; Length 14683;
Best Local Similarity 48.1%; Pred. No. 1.6e-23;
Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

```

```

QY 757 TACACTGGAATCCGATGAGAAAGATGACGTGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7348 TACCGGCTTCCGCTGGGATCCGCTCACTGACAGTGAAGGCGCAACATTTCTGATTAAC 7289
QY 817 GGGAAACCTGTCTTTTGAAGAGGCTTTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7288 CACAAACCGTTCTTACTTACTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7229
QY 877 GGCACCTTTTATCCATGATGATTAAGAGATTTCAACTTCTGAAGTGAATCAAGCGAAT 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7228 GGATTCGATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7169
QY 937 TCTTTCAGAGACTCTCACTATCTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7168 TCTTACCGTACTCGCATTTACCTTACGCTGAAGAGATGTGACTGGGAGAGATGAACAT 7109

```



```

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/723,624
/ FILING DATE: Concurrently Herewith
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Parker, David L.
/ REGISTRATION NUMBER: 32,165
/ REFERENCE/DOCKET NUMBER: BTIP:002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 512/418-3000
/ TELEFAX: 512/474-7577
/ INFORMATION FOR SEQ ID NO: 18:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3035 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-723-624-18

```

```

Query Match      6.6%; Score 110.8; DB 2; Length 3035;
Best Local Similarity 47.4%; Pred. No. 1.4e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

```

QY 757 TACACTCTGGACATCGGAATCGAAGCATGCTGGAGACAGAAAGGCTTATCTGAAC 816
DB 2016 TACCGGCTTGGGTCGGCATCCGGTCAGTGGAGTAAGGGGCAACAGTTCCGTAATAC 2075
QY 817 GGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAAATCCCGCTTGGGCGAG 876
DB 2076 CACAAACCTTCTACTTACTGCTGCTGCTCATGAAGATCCGACTTGGCGTGGCAAA 2135
QY 877 GGCACCTTTATTCATTGATGATTAAGAAGCTTCAACCTTCAAGTGGATCAACGGGAT 936
DB 2136 GGAATTCATACCTGCTGATGAGTGCAGACACGCACTTAATGACTGATGGGGCAAC 2195
QY 937 TCTTTGAGACCTCTCTCATCTCTTACAGTGAAGAGTGGCTGATCTTCCGACAGACTC 996
DB 2196 TCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATCTGACCTGGGCAATGAACAT 2255
QY 997 GGAATCTTGTGATGACGAAGAGCCCGCACGTTGGTATACAAAGT----- 1042
DB 2256 GGCATCGTGTGATGATGATAAACTGCTGCTGCTTAACTCTCTTTAGGCAATTGGT 2315
QY 1043 -----ACCACTACAAATCCGAGACT 1062
DB 2316 TTGGAAGCGGGCAACAGCCGAAAGAACTGTACAGCGAAAGGCACTCAACGGGAACT 2375
QY 1063 CAGAAAGATAGCAGAAAGACAATTAAGAAATGATGACAGACACAAAGACATCCCACT 1122
DB 2376 CAGCAAGCGCACTTAAGCGGATTAAGAGCTGATGCCGCTGACAAAACCAACCAAC 2435
QY 1123 GTGATCATGTGAGTGTGCGAAAGCAAGAGTCCCAACATCCAGACCGGAGGGTTTC 1182
DB 2436 GTGCGATGTGAGTGTGCTTCCAAAGAACCGGATACCCGCTCCCAAGGTGCAAGGGAATAT 2495
QY 1183 TTCAAAAGCCCTTATGAGCTGCAATGAAATGGAATGATGCAACGCCCGCTTGTCAATG 1242
DB 2496 TTGCGCCCACTGGCGGAAGCAACGCTAACTGACCCGACGCTCCGATCACTGCGCT 2555
QY 1243 AGCATGATGAGCAGCAGACAGAGAGAAACAAGAGAGTGGCGCTGAAGTACTTGCATATC 1302
DB 2556 AATGTATATTTCTGCGACGCTCAACAGATACCATCAAGGATCT-----CTTTATG 2609
QY 1303 GTCTGTGTAACAGGTAAGTACGCGCTGTATCATCTATCAGGAAAGATAGAAAGAGACTT 1362
DB 2610 CTGTGCTTAACCGTATTAACGATGTATGTCCAAAGCGGAGATTGAAACGCGCAGAG 2669
QY 1363 CAAGCTCTGGAAGAAACATAGAAAGCTCTATGCAAGGCAACAAAGCCATCTTTGTC 1422

```

```

DB 2670 AAGGTAAGTGAAGAAAGAACTTGTGCGCTGGCAGAGAAAGTCAACAGCCGATTTATCATC 2729
QY 1423 ACAGAAATTCGTCGCGACCGGATAGCTGGCATTCACATGATCCACTTCAATGTTCTCC 1482
DB 2730 ACCGAATAGCGGCGTGAATACGTTAGCCGGGCTGCACCTCAATATACCGACATGTGAGAT 2789
QY 1483 GAAGAGTACCAAGCAGAGCTCTTGAAGAAAGATCAGGCTCTTTGAAAAAAGACTAC 1542
DB 2790 GAAGAGTACAGTGTGATGCTGATGCTGATATGATACCGGCTTTGATGCTGACGCCG 2849
QY 1543 ATCATCGAAACACAGTGTGGGCTTTGAGATTTTAAGACTCTCAAGATGAGAAAGA 1602
DB 2850 GTGTCGGTGAACAGATGATGATTTCCCGATTTTGGCACTCCGACAGCATATTTGCC 2909
QY 1603 CCCATTCTCAACCAAGGCTTTTCAACAGAGACAGAACCAACCACTGCTGCTCAT 1662
DB 2910 GTTGGCGGTAAACAAAGAGGATCTTCACTGCGCAGCCGAAACCGAAGTGGCGGCTTT 2969
QY 1663 GTACTGAGAAAGCTGTGAGTG 1684
DB 2970 CTGTCGCAAAAACGCTGACTG 2991

```

```

RESULT 11
US-08-630-820-5
/ Sequence 5, Application US/08630820
/ Patent No. 6008023
/ GENERAL INFORMATION:
/ APPLICANT: OPPER, Martin
/ APPLICANT: BOSSLET, Klaus
/ APPLICANT: CZECH, Joerg
/ TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
/ TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Foley & Lardner
/ STREET: 3000 K Street, N.W., Suite 500
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20007-5109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/630, 820
/ FILING DATE: 10-APR-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: DE 19513676.4
/ FILING DATE: 11-APR-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: GRANADOS, Patricia D.
/ REGISTRATION NUMBER: 33,683
/ REFERENCE/DOCKET NUMBER: 18748/306
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 672-5300
/ TELEFAX: (202) 672-5399
/ TELEX: 904136
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3169 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: CDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Enterobacteriaceae: Escherichia coli

```

STRAIN: PRJ210
 IMMEDIATE SOURCE:
 CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 3..641
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 666..3162
 US-08-630-820-5

Query Match 6.6%; Score 110.8; DB 3; Length 3169;
 Best Local Similarity 47.4%; Pred. No. 1.5e-23;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TACACTCTGGACATGGAGTACGAACGATCAGCTGGAGAGAGAGGCTATCTGAAC 816
 DB 2145 TACCCGCTCCGCGCGCATCCGCTCAGTGGACGTGAAGGCGCAACAGTTCTGTATTAAC 2204
 QY 817 GGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGCGAG 876
 DB 2205 CACAAACGTTCTACTCTTACTGCTTTGGTGTGCTGATGAAGATGCGGACTTACGTGGCAA 2264
 QY 877 GGCACCTTTTATCATTTGATGATTAAGACTTCAACTTTTGAAGTGATCAACCGCAAT 936
 DB 2265 GGATTCGATTAACGTGCTGATGGTGACGACCAACGCAATTATGACTGATGGGCGCAAC 2324
 QY 937 TCTTTACAGACCTTCACATATCTTACAGTGAAGAGTGGCTGATCTTGGCGACAGATC 996
 DB 2325 TCTTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGGAGATGAACAT 2384
 QY 997 GGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTGATCAAGGT----- 1042
 DB 2385 GGCATGCTGTGATTTGATGAATCTGCTGTGCTTAACTCTCTTTAGGCAATGGT 2444
 QY 1043 -----ACCACTACAACTCCGAGACT 1062
 DB 2445 TTCGAAGCGGCAACAAAGCCGAAGAACTGTACAGCGAAGAGCACTCAACGGGAAAACT 2504
 QY 1063 CAGAGATAGCAAGAGCAAAATTAAGAAATGATGATGACAGACACAAGACATCCCACT 1122
 DB 2505 CAGCAAGCGCACTTACAGCGCATTTAAAGCTGATAGCGCGTGAACAAAACCAACCAAGC 2564
 QY 1123 GTGATCATGTGAGTGTGCGAAGCAAGACAGAGTCCAACTCAGACCGGAGGGTTTC 1182
 DB 2565 GTGGATGTGAGTATGCTCAACGAACCGGATACCCGTCGCGAAGGTGCAAGGAATAT 2624
 QY 1183 TTCAAAACCTTTTATGAGACTGCGCAATGAATGATCGAAACGCGCCGTTGTCAATG 1242
 DB 2625 TTCGGGCCACTGGCGGAAGCAACGCGTAACTGACCCGACGCTCCGATCACTGCGTTC 2684
 QY 1243 AACATGATGAGACGACCAAGAGAGAAACAAGACGTGGCGCTGAAGTACTTGACATC 1302
 DB 2685 AATGTATATTTTTCGCAACCTCAACCGATCAATCAGCATCT-----CTTTATGTG 2738
 QY 1303 GTCTGTGTGAACAGGACTACTACGCTGTGTATCATCTATCAGGGAAGATAGAAAGAACTT 1362
 DB 2739 CTGTGCTGAACGTTATTAACGAGATGTATGTCCAAAGCGGAGATTTGGAACGGCAAG 2798
 QY 1363 CAACTCTTGAAGAAACATAGAGAGCTGTATGCAAGGACAGAAAGCCATCTTTGTC 1422
 DB 2799 AAGGTACTGGAAGAAAGAACTTCTGCTGCGAGGAGAACTGCATCAGCGATTTATCATC 2858
 QY 1423 ACAGATTGCTGGGAGCGGATAGCTGGCATCCATACATCCACTCAATATGTTTCC 1482
 DB 2859 ACCGAATTACGGCGGTGATAGCTTTAGCGGCTGCACTCAATGATACCGCAATGTGAGT 2918
 QY 1483 GAAGATACCAAGCAGAGCTGTGTAAGAAAGACATAGGCTCTTTTGAAGAAAGACTAC 1542
 DB 2919 GAAGATACCAAGTGTGATGCTGATATGATACCGGCTCTTTGATGCGCTGACGCGC 2978
 QY 1543 ATCATGGAACACGCTGTGGGCTTTGCAATTTTAAAGATCTTCAGAAATGTGAAGA 1602

DB 2979 GTCCGCGTGAACAGTATGAAATTTGCGCATTTTGGCACTCGCAAGCATATTTGGC 3038
 QY 1603 CCATTCACACCAAGAGGCTTTTCAACAGACAGACCAACCAACTGTTGCTCAT 1662
 DB 3039 GTGGCGGTAAACAAGAAAGGATCTTCACTGCAACGCAACGAAAGTGGCGGCTTTT 3098
 QY 1663 GTACTGAGAGAGCTGTGAGTGTG 1684
 DB 3099 CTGCTGCAAAACCTGTGACTG 3120

RESULT 12
 US-09-273-453-5
 Sequence 5, Application US/09273453
 Patent No. 6602688

GENERAL INFORMATION:

APPLICANT: OPPER, Martin
 BOSSLET, Klaus
 CZECH, Joerg

TITLE OF INVENTION: CYTOSOLIC EXPRESSION OF ANTIBODIES,
 ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES

IN E. COLI

NUMBER OF SEQUENCES: 7

RESPONSE ADDRESS:
 ADDRESS: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,453

FILING DATE: 22-Mar-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/630,820

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: GRANADOS, Patricia D.

REGISTRATION NUMBER: 33,683

REFERENCE/DOCKET NUMBER: 18748/306

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3169 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterobacteriaceae: Escherichia coli

STRAIN: PRJ210

IMMEDIATE SOURCE:

CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc

FEATURE:

NAME/KEY: CDS

LOCATION: 666..3162

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-273-453-5

Query Match 6.6%; Score 110.8; DB 4; Length 3169;

Best Local Similarity 47.4%; Pred. No. 1.5e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

QY 757 TACACTGGAACATCGGAATCAGACGATCGGACGAGAGAGGCTCTATCTGAAC 816
    |||
DB 2145 TACCCGCTTCGCGTCCGATCCGATCGATGAGAGAGGCGAACAAGTCTGATTAAC 2204
    |||
QY 817 GGGAAACCTGCTTTTGAAGGCTTTGAAAGACAGAGGAAATTCCTCGTTGGGCGAG 876
    |||
DB 2205 CACAAACGTTCTACTTACTGCTTGGTGGTCTGATGAGATGCGAGCTTACCTGGCAA 2264
    |||
QY 877 GGCACCTTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
    |||
DB 2265 GGATTCGATTAACGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2324
    |||
QY 937 TCTTTGAGACCTCTCTCACTATCTCTTACAGTGAAGAGTGGTGGATCTTGGCGACAATC 996
    |||
DB 2325 TCGTACCGTACCTTCGATTAACCTTACGCTGAAGAGATGCTGACCTGGGCGAGATGAACAT 2384
    |||
QY 997 GGAATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1042
    |||
DB 2385 GGCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2444
    |||
QY 1043 -----ACCACTACATCCCGAGACT 1062
    |||
DB 2445 TTGGAAGCGGGCAACAAGCCGAAAGAACTGTACAGGGAAGGCGAGTCAACGGGGAACT 2504
    |||
QY 1063 CAGAAATAGCAGAAAGCAACATTAAGAGATGATGACAGACACAAAGCAATCCCACT 1122
    |||
DB 2505 CAGCAAGCGCACTTACAGGCGATTAAGAGCTATAGCGGTGACAAACCAACCAACG 2564
    |||
QY 1123 GTGATCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1182
    |||
DB 2565 GTGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2624
    |||
QY 1183 TTCMAAGCCCTTATGAGACTGCCAATGAATGATGATGATGATGATGATGATGATGATG 1242
    |||
DB 2625 TTGCGGCGCACTGGCGGAGAGCAAGCGTAACTGACCCGAGCTGCTCCATCTCCGCTC 2684
    |||
QY 1243 AGCATATGAGACGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302
    |||
DB 2685 AATGATATGTTCTGCGACGCTCACACGATACCATCAGCGATCT-----CTTTGATG 2738
    |||
QY 1303 GTCTGTGTAACAGGATTAACGCTGATCATCTATCAGGAGAGATGAAGAGAGACTT 1362
    |||
DB 2739 CTGTGCTTAACCGTTATTAAGGATGATGATGATGATGATGATGATGATGATGATGAT 2798
    |||
QY 1363 CAAAGCTCTGGAAGAAAGACATAGAAAGAGCTATGCAAGGACAGAAAGCCATCTTTGTC 1422
    |||
DB 2799 AAGGATCTGGAAGAAAGAACTTCTGCGCTGCGAGAGAAAGTGCATCAGCGATTAATC 2858
    |||
QY 1423 ACAGAAATTCGTCGCGAGCGGATAGCTGATGATGATGATGATGATGATGATGATGATG 1482
    |||
DB 2859 ACCGAATACGCGCGTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2918
    |||
QY 1483 GAAAGATACCAAGACAGAGCTCGTTGAAAAGAGATGATGATGATGATGATGATGATGAT 1542
    |||
DB 2919 GAAAGATATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2978
    |||
QY 1543 ATCATCGGAACACAGCTGTGGGCTTTGAGATTTTAAGATCTCTGAGATGAGAGAGA 1602
    |||
DB 2979 GTGCTGCGTGAACAGGATTAAGAAATTTGCGCGATTTTTCGACCTCGGAAGCAATTTG 3038
    |||
QY 1603 CCATTTCTCAACCAAGAGGTTTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662
    |||
DB 3039 GTTGGCGGTAAACAAAGAGAGAGATCTTCACTCGGAGCCGAAACGAAAGTGGGCGGTT 3098
    |||
QY 1663 GTAATGAGAGAGAGCTGTGAGAGT 1684
    |||
DB 3099 CTGCTGCAAAAACGCTGAGCTG 3120
    |||

```

RESULT 13

US-08-723-624-19
Sequence 19, Application US/08723624
Patent No. 5861277

GENERAL INFORMATION:
APPLICANT: Rose, Alan B.
APPLICANT: Last, Robert L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,624
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: BTIP:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 3824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-723-624-19

Query Match 6.6%; Score 110.8; DB 2; Length 3824;
Best Local Similarity 47.4%; Pred. No. 1.7e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

QY 757 TACACTTGGACATGGAATCAGAAAGATCACTGAGAGAGAGAGAGAGAGAGCTTATCTGAAC 816
    |||
DB 2805 TACCCGCTTCGCGTCCGATCCGATCGATGAGAGAGAGAGAGAGAGAGCTTATTAAC 2864
    |||
QY 817 GGGAAACCTGCTTTTGAAGGCTTTGAAAGACAGAGAAATTCCTCGTTGGGCGAG 876
    |||
DB 2865 CACAAACGTTCTACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2924
    |||
QY 877 GGCACCTTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
    |||
DB 2925 GGATTCGATTAACGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2984
    |||
QY 937 TCTTTGAGACCTCTCACTATCTTACAGTGAAGAGTGGATCTTGGCGAGACTC 996
    |||
DB 2985 TCTTACCGTACCTGCTGATTAACCTTACGCTGAAGAGATGCTGAGCTGGGCGAGATGA 3044
    |||
QY 997 GGAATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1042
    |||
DB 3045 GGCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3104
    |||
QY 1043 -----ACCACTACATCCCGAGACT 1062
    |||
DB 3105 TTGGAAGCGGGCAACAAGCCGAAAGAACTGTACAGCGAAGGCGAGTCAACGGGAAACT 3164
    |||
QY 1063 CAGAAATAGCAGAAAGCAACATTAAGAGATGATGACAGACACAAAGCAATCCCACT 1122
    |||
DB 3165 CAGCAAGCGCACTTACAGGCGATTAAGAGAGAGATGATGATGATGATGATGATGATGAT 3224
    |||
QY 1123 GTGATCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1182
    |||

```



```

RESULT 15
US-09-893-525-41
/ Sequence 41, Application US/09893525
/ Patent No. 6753167
/ GENERAL INFORMATION:
/ APPLICANT: Moloney, Maurice M.
/ APPLICANT: Van Rooijen, Gijb
/ TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
/ FILE REFERENCE: 9369-172
/ CURRENT APPLICATION NUMBER: US/09/893,525
/ CURRENT FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 09/210,843
/ PRIOR FILING DATE: 1998-12-15
/ PRIOR APPLICATION NUMBER: US 08/846,021
/ PRIOR FILING DATE: 1997-04-25
/ PRIOR APPLICATION NUMBER: US 08/366,783
/ PRIOR FILING DATE: 1994-12-30
/ PRIOR APPLICATION NUMBER: US 08/142,418
/ PRIOR FILING DATE: 1993-11-16
/ PRIOR APPLICATION NUMBER: US 07/659,835
/ PRIOR FILING DATE: 1991-02-22
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 41
/ LENGTH: 5390
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: phas-calco-GUS-phas
/ NAME/KEY: CDS
/ LOCATION: (1548)..(4097)
/ OTHER INFORMATION:
US-09-893-525-41

```

Query Match 6.6%; Score 110.8; DB 4; Length 5390;
 Best Local Similarity 47.4%; Pred. No. 2.1e-23;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

QY 757 TACACTCTGAGATCGGAATCAGAACGATCAGCTGGACGAGAAGAGCTTATCTGAAAC 816
DB 3081 TACCGGCTTCGGGTGGGATCCGGTCAGTGGCAGTGAAGGCCAACAGTTCCGATTAAC 3140
QY 817 GGGAAACCTGTCTTTTGAAGGGCTTTGAAAGACGAGAAATCCCGTTCTGGGGCAG 876
DB 3141 CACAAACCGTCTACTTACTGCTTGGCTTGGTCGTCATGAAGATCGGACTTACGTGGCAA 3200
QY 877 GGCACCTTTATCCATGATGATGAATAAAGCTTCAACCTTCAAGTGAATCAACGGAT 936
DB 3201 GGAATTCGATTAACGTGCTGATGGTGCACGACCAAGCATTAATGACTGATGGGCGCAAC 3260
QY 937 TCTTTCAGAGCTCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTCCGACAGACTC 996
DB 3261 TCTTACCGTACTCGATTAACCTTACGCTGAAGAGATGCTGAGCTGGCAGATGAACAT 3320
QY 997 GGAATCTTGTGATAGCAAGAGCCCGCACGTTGTATCAAAAGT----- 1042
DB 3321 GGCATCGTGTGATGATGATAAATGCTGCTGCTTTCGCTCTTTAGGCATTGGT 3380
QY 1043 -----ACCATCAATCCGAGACT 1062
DB 3381 TTCCGAGCGGGCAACAGCCGAAAGAACTGTACAGCGAAGAGGCACTCAACGGGAAACT 3440
QY 1063 CAGAAATAGCAGAAAGAACATTAAGAAAGATGATGACAGACACAAAGACATCCCACT 1122
DB 3441 CAGCAGCGCCTTACAGGCGATTAAGAGCTGATAGCCGCTGACAAAACCAACCAAGC 3500
QY 1123 GTGATCATGTGAGTGTGCGAAAGCAACGAGTCCAACTCAGACCGGAGGGTTTC 1182
DB 3501 GTGGTGTATGTGATGATTCGCAACGAAACCGGATACCCGTCGCAAGGTGACAGGGAAATAT 3560
QY 1183 TTCAAAAGCCCTTTATAGATGTCGCAATGAATGATCGAAACAGCCCGCTTGTCAATGTG 1242

```

```

DB 3561 TTCCGCGCATCTGGCGGAAGCAACGGTAAACTGCACCCGAGCGCTCCGATCATCTGCCGC 3620
QY 1243 AGCATATGAGACGCAACGAGAGAAACAGAGACGTGGCCGCTGAAGTACTTGCATATC 1302
DB 3621 AATGTAAATGTTCTGGAGCGCTCACACCGATACATCAGCGATCT-----CTTGTATGTG 3674
QY 1303 GTCTGTGGAACAGTACTACGCGCTGGTACATCTATCAGGGGAAGATGAAGAAGACTT 1362
DB 3675 CTGTCCCTGAACCGTTATTTACGATGTATGTTCAAACGCGGATTTGAAACGGCAGAG 3734
QY 1363 CAAGCTGTGAAAAAGACATAGAAAGCTTATGCAAGGCAAGAAAGCCCATCTTTTGTG 1422
DB 3735 AAGTACTGTGAAAAAGAACTTGTGCGCTGGCAGAGAAATCTCATCAGCCGATTTATCATC 3794
QY 1423 ACAGATTCGTGGCGGACGCGATAGCTGATCATCTACATCATCTTCAAAATGTTCTCC 1482
DB 3795 ACCGAATACGGCGGTGATACGTTACCCGGGCTGCATCAATGTACACCGACATGTGAGT 3854
QY 1483 GAAGAGTACCAAGCAGAGCTGTTGAAAAAGCATCAGGCTCCTTTGAAAAAAGACTAC 1542
DB 3855 GAAGGATACAGTGTGATAGCTGATATGATACCCGCGTCTTGTATGCGCTCAGCGCC 3914
QY 1543 ATCATCGGAACACAGGTGGGCTTTGCAATTTTAAGACTCTCAGAAATGTGAGAAGA 1602
DB 3915 GTGCTCGGTGAACAGATATGAATTTCCCGCATTTTGGGACCTCGCAAGCATATTGGCC 3974
QY 1603 CCATTTCAACCAAGGGTGTTTTCAACAGAGCAGACAAACCAAACTCGTGTCTCAT 1662
DB 3975 GTTGGCGTAAACAAAGAGGATCTTCACTCGCGACCGCAAAACGAAATCGGCGGCTTT 4034
QY 1663 GTACTGAGAAGACTGTGAGTGC 1684
DB 4035 CTGCTCAAAAAACGCTGACATG 4056

```

Search completed: September 1, 2005, 19:13:25
 Job time : 340 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 17:43:49 : Search time 1163 Seconds
(without alignments)
9510.845 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atggttaagccgcgaacgaaa.....gaagactgtgagtgagctt 1689

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 733684 seqs, 327445616 residues

Total number of hits satisfying chosen parameters: 14677368

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	17	US-10-364-649-14
2	1688	99.9	1689	17	US-10-369-493-26682
3	175.4	10.4	1926	22	US-10-757-093-1
4	174	10.3	2169	14	US-10-136-841-5
5	174	10.3	2169	17	US-10-272-483A-5
6	174	10.3	2169	17	US-10-272-483A-5
7	172.4	10.2	1956	17	US-10-421-175-1

8	172.4	10.2	2191	17	US-10-388-360-343	Sequence 343, App
9	172.4	10.2	2191	17	US-10-429-802-16	Sequence 16, Appl
10	172.4	10.2	2191	17	US-10-430-503-7	Sequence 7, Appl
11	172.4	10.2	2191	20	US-10-335-053-51	Sequence 51, Appl
12	166.8	9.9	2472	17	US-10-388-934-225	Sequence 225, App
13	159.6	9.4	1947	17	US-10-421-175-3	Sequence 3, Appl
14	141.6	8.4	1905	22	US-10-757-093-3	Sequence 3, Appl
15	141.4	8.4	1987	16	US-10-120-145-7	Sequence 7, Appl
16	141.4	8.4	3651	16	US-10-322-656-45	Sequence 45, Appl
17	141.4	8.4	4084	16	US-10-322-656-48	Sequence 48, Appl
18	137.6	8.1	1905	22	US-10-757-093-5	Sequence 5, Appl
19	135.4	8.0	1935	22	US-10-757-093-9	Sequence 9, Appl
20	130.4	7.7	1888	17	US-10-364-649-27	Sequence 27, Appl
21	126.2	7.5	1939	22	US-10-757-093-7	Sequence 7, Appl
22	114	6.7	3451	15	US-10-161-403-108	Sequence 108, App
23	114	6.7	3451	20	US-10-161-408-20	Sequence 20, Appl
24	114	6.7	3451	24	US-11-006-076-108	Sequence 108, App
25	114	6.7	14627	15	US-10-161-403-109	Sequence 109, App
26	114	6.7	14627	20	US-10-161-408-21	Sequence 21, Appl
27	114	6.7	14627	24	US-11-006-076-109	Sequence 109, App
28	112.4	6.7	1812	24	US-11-000-863-13	Sequence 13, Appl
29	112	6.6	14683	13	US-10-109-853-1	Sequence 1, Appl
30	112	6.6	14683	19	US-10-817-950-1	Sequence 1, Appl
31	110.8	6.6	1809	19	US-10-432-777-16	Sequence 16, Appl
32	110.8	6.6	1812	15	US-10-161-403-105	Sequence 105, App
33	110.8	6.6	1812	17	US-10-369-493-24529	Sequence 24529, A
34	110.8	6.6	1812	20	US-10-161-408-16	Sequence 16, Appl
35	110.8	6.6	1812	20	US-10-149-533A-33	Sequence 33, Appl
36	110.8	6.6	1812	24	US-11-006-076-105	Sequence 105, App
37	110.8	6.6	2001	18	US-10-239-907A-38	Sequence 38, Appl
38	110.8	6.6	2001	19	US-10-332-406A-21	Sequence 21, Appl
39	110.8	6.6	2001	19	US-10-800-161-29	Sequence 29, Appl
40	110.8	6.6	2186	20	US-10-755-328-2	Sequence 2, Appl
41	110.8	6.6	2725	18	US-10-239-907A-43	Sequence 43, Appl
42	110.8	6.6	2730	18	US-10-239-907A-41	Sequence 41, Appl
43	110.8	6.6	3169	21	US-10-632-815-5	Sequence 5, Appl
44	110.8	6.6	4072	18	US-10-239-907A-49	Sequence 49, Appl
45	110.8	6.6	4341	18	US-10-239-907A-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-10-364-649-14
; Sequence 14, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1689)
; OTHER INFORMATION: n = A,T,C or G
US-10-364-649-14
Query Match 99.9%; Score 1688; DB 17; Length 1689;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTAAAGCCGCAAGAAACAAGAGAGATTATCTTATCTTGAATGGAGTTGGAT 60
 DB 1 ATGGTAAAGCCGCAAGAAACAAGAGAGATTATCTTATCTTGAATGGAGTTGGAT 60
 QY 61 CTGGAAGTAACGACAAAGACAGACCAATCGCCGTTCTTGAAAGCTGAATGAGAGTAC 120
 DB 61 CTGGAAGTAACGACAAAGACAGACCAATCGCCGTTCTTGAAAGCTGAATGAGAGTAC 120
 QY 121 CAGGATCTGTGCTACGAAAGAGACCTTTCACCTTACAAAACAACCTTCTAGCTTCCGAAG 180
 DB 121 CAGGATCTGTGCTACGAAAGAGACCTTTCACCTTACAAAACAACCTTCTAGCTTCCGAAG 180
 QY 181 NAACTTCAAAAAACATCAGACTTTTACTTTGCTGGGTGAACAAGGCTGGAGATG 240
 DB 181 NAACTTCAAAAAACATCAGACTTTTACTTTGCTGGGTGAACAAGGCTGGAGATG 240
 QY 241 TTCTCTCAACGAGAGAAAGTGGAGAGAAATCAATTGAATACCTTCCCTTCCAGTAGAT 300
 DB 241 TTCTCTCAACGAGAGAAAGTGGAGAGAAATCAATTGAATACCTTCCCTTCCAGTAGAT 300
 QY 301 GTGACGGGGAAAGTGAATCCCGAGAGAAAGAACTGAGGTGTTGTAAGACAGATTG 360
 DB 301 GTGACGGGGAAAGTGAATCCCGAGAGAAAGAACTGAGGTGTTGTAAGACAGATTG 360
 QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGGATTTT 420
 DB 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGGATTTT 420
 QY 421 GGAAGTTTTCAACCTGCAAACTTTCGACTTTCCTCCCTCAAGTGAATCAAGGCTGTT 480
 DB 421 GGAAGTTTTCAACCTGCAAACTTTCGACTTTCCTCCCTCAAGTGAATCAAGGCTGTT 480
 QY 481 CTGATAGATTTCAACAGACCGAGAGATTAATCTGAGATCTGGGTGACACGAGTAGT 540
 DB 481 CTGATAGATTTCAACAGACCGAGAGATTAATCTGAGATCTGGGTGACACGAGTAGT 540
 QY 541 GAACCGAGAGAAACTTGGAAAGTGAAGTGAAGATGAAGTCTCAGAAAGAGGAGT 600
 DB 541 GAACCGAGAGAAACTTGGAAAGTGAAGTGAAGTGAAGTCTCAGAAAGAGGAGT 600
 QY 601 GGAACGAGAGATGACGATCAAACTTGGAGAGAAAGAAAGATTAGAACATCCAACGA 660
 DB 601 GGAACGAGAGATGACGATCAAACTTGGAGAGAAAGAAAGATTAGAACATCCAACGA 660
 QY 661 TTCTGTGAAGGGAGATTCTCTCCGAAACGCCAGGTTTGGAGCTCGAAGATCCATAT 720
 DB 661 TTCTGTGAAGGGAGATTCTCTCCGAAACGCCAGGTTTGGAGCTCGAAGATCCATAT 720
 QY 721 CTCTTATCTCTCAAGGTGGAACCTTGAAGAAAGACAGATCACTCTGGAATCGGAATCGA 780
 DB 721 CTCTTATCTCTCAAGGTGGAACCTTGAAGAAAGACAGATCACTCTGGAATCGGAATCGA 780
 QY 781 ACGATCAGCTGGGACGAGAAAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGC 840
 DB 781 ACGATCAGCTGGGACGAGAAAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGC 840
 QY 841 TTTGGAAGACAGAGAAATTTCCCGTTCTGGGGACGGGACCTTTTATCTGATGATGATA 900
 DB 841 TTTGGAAGACAGAGAAATTTCCCGTTCTGGGGACGGGACCTTTTATCTGATGATGATA 900
 QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTTCAGAGACCTCTCATATCCT 960
 DB 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTTCAGAGACCTCTCATATCCT 960
 QY 961 TACAGTGAAGAGTGGCTGATCTTTCGACAGACTTCGGAATCTTGTGTATGACGAAGCC 1020
 DB 961 TACAGTGAAGAGTGGCTGATCTTTCGACAGACTTCGGAATCTTGTGTATGACGAAGCC 1020
 QY 1021 CCGCAGCTTGTGATCAAGAGTACCACTTCAATCCGAGACTCAGAAAGTACAGAAAGC 1080
 DB 1021 CCGCAGCTTGTGATCAAGAGTACCACTTCAATCCGAGACTCAGAAAGTACAGAAAGC 1080

QY 1081 AACATGAAGAAATGATGACAGACAGAAAGCAATCCCAAGTGTATGATGAGTGTG 1140
 DB 1081 AACATGAAGAAATGATGACAGACAGAAAGCAATCCCAAGTGTATGATGAGTGTG 1140
 QY 1141 GCGAAGCAAGAGTCCCAACCATTCAGACGCGAGAGGTTTCTTCAAGCCCTTTATGAG 1200
 DB 1141 GCGAAGCAAGAGTCCCAACCATTCAGACGCGAGAGGTTTCTTCAAGCCCTTTATGAG 1200
 QY 1201 ACTGCCAATGAAATGATTCGAACACGCCCCCTTGTCTATGTGTAGCATATGAGCGCACCA 1260
 DB 1201 ACTGCCAATGAAATGATTCGAACACGCCCCCTTGTCTATGTGTAGCATATGAGCGCACCA 1260
 QY 1261 GACGAGAGAACAGAGAGCTGGCGCTGAAGTACTTCGACATCGTGTGTGAACGAGTAC 1320
 DB 1261 GACGAGAGAACAGAGAGCTGGCGCTGAAGTACTTCGACATCGTGTGTGAACGAGTAC 1320
 QY 1321 TACGGCTGTATCATCTATCAAGGAAAGATAGAAAGGACTTCAAGCTCTGAAAAAAGAC 1380
 DB 1321 TACGGCTGTATCATCTATCAAGGAAAGATAGAAAGGACTTCAAGCTCTGAAAAAAGAC 1380
 QY 1381 ATAGAAAGCTCTATGCAAGGACACAGAAAGCCATCTTGTCAAGAAATTTGGTGGCGAC 1440
 DB 1381 ATAGAAAGCTCTATGCAAGGACACAGAAAGCCATCTTGTCAAGAAATTTGGTGGCGAC 1440
 QY 1441 GCGATAGCTGGCATTCACATCAAGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
 DB 1441 GCGATAGCTGGCATTCACATCAAGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
 QY 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTTGAAGAAAGCTACATCATCGAAACACAGTG 1560
 DB 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTTGAAGAAAGCTACATCATCGAAACACAGTG 1560
 QY 1561 TGGGCTTTGCAAGTTTAAAGCTCCGCAAGATGGAAGAACCCATCTTCAACACAGAG 1620
 DB 1561 TGGGCTTTGCAAGTTTAAAGCTCCGCAAGATGGAAGAACCCATCTTCAACACAGAG 1620
 QY 1621 GGTGTTTCAACAAGAGACAGAACCCAACTCGTGTGCTCACTGATGAGAAAGCTGTGG 1680
 DB 1621 GGTGTTTCAACAAGAGACAGAACCCAACTCGTGTGCTCACTGATGAGAAAGCTGTGG 1680
 QY 1681 AGTGAAGTT 1689
 DB 1681 AGTGAAGTT 1689

RESULT 2
 US-10-369-493-26682
 ; Sequence 26682, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369, 493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 26682
 ; LENGTH: 1692
 ; TYPE: DNA
 ; ORGANISM: Thermotoga maritima
 US-10-369-493-26682

Query Match 99.9%; Score 1688; DB 17; Length 1692;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

QY 1 ATGTAAGACCGCAAGAAACAAGAGATTTATTCTTATCTTGAATGGAGTTGGAAAT 60
   |||||
Db 1 ATGTAAGACCGCAAGAAACAAGAGATTTATTCTTATCTTGAATGGAGTTGGAAAT 60
QY 61 CTTGAAGTAACGACAAAGACAGACCAATCGCGTTCTTGAAGTGTGAATGACAGTAC 120
   |||||
Db 61 CTTGAAGTAACGACAAAGACAGACCAATCGCGTTCTTGAAGTGTGAATGACAGTAC 120
QY 121 CAGGATCTGTGCTACGAGAAAGAGACCTTCACTTACAAAACAACCTTTCTACGTTCCGAAG 180
   |||||
Db 121 CAGGATCTGTGCTACGAGAAAGAGACCTTCACTTACAAAACAACCTTTCTACGTTCCGAAG 180
QY 181 NAACTTCAAAAAACATCAGACTTAACTTTGCTGCGGTGAACAACGACTGCGAGTTC 240
   |||||
Db 181 NAACTTCAAAAAACATCAGACTTAACTTTGCTGCGGTGAACAACGACTGCGAGTTC 240
QY 241 TTCTCTCAACGAGAGAAAGTGGAGAGAAATCATTGAATACCTTCCCTTCAAGTATGAT 300
   |||||
Db 241 TTCTCTCAACGAGAGAAAGTGGAGAGAAATCATTGAATACCTTCCCTTCAAGTATGAT 300
QY 301 GTGACGGGGAAGTGAATCCGAGAGAACGAATCTAGGGGTGTTGTAAGACAGATTG 360
   |||||
Db 301 GTGACGGGGAAGTGAATCCGAGAGAACGAATCTAGGGGTGTTGTAAGACAGATTG 360
QY 361 AAAGTGGAGAGATTTCTCTGAGATTCAGACAGGGGCACTCACCCGTTGGGATTTT 420
   |||||
Db 361 AAAGTGGAGAGATTTCTCTGAGATTCAGACAGGGGCACTCACCCGTTGGGATTTT 420
QY 421 GGAAGTTTTCACCTGCAAACTTCGACTTTTCCCTTACGTTGGAATCAATAGGCTGT 480
   |||||
Db 421 GGAAGTTTTCACCTGCAAACTTCGACTTTTCCCTTACGTTGGAATCAATAGGCTGT 480
QY 481 CTGATAGACTTCACAGACACGCGAGATATCTGACATCTGGGTGACACGAGTGTCT 540
   |||||
Db 481 CTGATAGACTTCACAGACACGCGAGATATCTGACATCTGGGTGACACGAGTGTCT 540
QY 541 GAAACGGAGAGAACTTGGAAAGTGAAGTGAAGTCAAGTCAAGAAAGGCGTG 600
   |||||
Db 541 GAAACGGAGAGAACTTGGAAAGTGAAGTGAAGTCAAGTCAAGAAAGGCGTG 600
QY 601 GGAAGAGATGACGATCAAACTTGGAGAGAGAAAGAAAGATTAGAACATCCACAGA 660
   |||||
Db 601 GGAAGAGATGACGATCAAACTTGGAGAGAGAAAGAAAGATTAGAACATCCACAGA 660
QY 661 TTGCTGGAAGGGAGTTTCACTCTCGAAAACGCCAGTTTGAAGCTCGAATCCATAT 720
   |||||
Db 661 TTGCTGGAAGGGAGTTTCACTCTCGAAAACGCCAGTTTGAAGCTCGAATCCATAT 720
QY 721 CTTTATCTCTCAAGGTGGAATTTGAAAAAGAGATTAACCTTGAACATCGGAATCA 780
   |||||
Db 721 CTTTATCTCTCAAGGTGGAATTTGAAAAAGAGATTAACCTTGAACATCGGAATCA 780
QY 781 ACGATCAGCTGGGACAGAAAGAGGCTCTATCTGAAAGGGAAACCTGTCTTTTGAAGGC 840
   |||||
Db 781 ACGATCAGCTGGGACAGAAAGAGGCTCTATCTGAAAGGGAAACCTGTCTTTTGAAGGC 840
QY 841 TTTGGAAGACGAGAGAAATTCGCCGTTCTGGGGCAGGGACCTTTTATCATTTGATAT 900
   |||||
Db 841 TTTGGAAGACGAGAGAAATTCGCCGTTCTGGGGCAGGGACCTTTTATCATTTGATAT 900
QY 901 AAAAGACTTCAACCTTTGAAAGTGAATCAACGGAATTTCTTCAAGACCTTCACTATCT 960
   |||||
Db 901 AAAAGACTTCAACCTTTGAAAGTGAATCAACGGAATTTCTTCAAGACCTTCACTATCT 960
QY 961 TACAGTGAAGAGTGTGATCTTGGCAGACAGACTCGGAATCTTGTGTAAGCAAGGC 1020
   |||||
Db 961 TACAGTGAAGAGTGTGATCTTGGCAGACAGACTCGGAATCTTGTGTAAGCAAGGC 1020
QY 1021 CCGCAGCTTGTGATCAAAAGTACCAATCCGAGACTCAGAAAGTAAAGCAAGAC 1080
   |||||
Db 1021 CCGCAGCTTGTGATCAAAAGTACCAATCCGAGACTCAGAAAGTAAAGCAAGAC 1080
QY 1081 AACATGAAGAAATGATCGACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
   |||||

```

```

Db 1081 AACATGAAGAAATGATGACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
   |||||
QY 1141 GCGAAGCAACAGAGTCCAAACCATCCAGACGGGAGGTTTCTTCAAAAGCCCTTATGAG 1200
   |||||
Db 1141 GCGAAGCAACAGAGTCCAAACCATCCAGACGGGAGGTTTCTTCAAAAGCCCTTATGAG 1200
QY 1201 ACTGCCAATGAATGATGTGAACAGGCCCGTTGTTCATGTGTGACATGTGACACCA 1260
   |||||
Db 1201 ACTGCCAATGAATGATGTGAACAGGCCCGTTGTTCATGTGTGACATGTGACACCA 1260
QY 1261 GACGAGAGAAACAAGAGACGTGGCGCTGAAGTACTTCAATGTCGTGTGAAACAGTAC 1320
   |||||
Db 1261 GACGAGAGAAACAAGAGACGTGGCGCTGAAGTACTTCAATGTCGTGTGAAACAGTAC 1320
QY 1321 TACGGCTGTACATCTATCAGGAGAAAGATAGAAGAGACTTCAAGCTTGAAGAAAGAC 1380
   |||||
Db 1321 TACGGCTGTACATCTATCAGGAGAAAGATAGAAGAGACTTCAAGCTTGAAGAAAGAC 1380
QY 1381 ATAGAAGACTCTATGCAAGGACACAGAAAGCCCATCTTTGTCAACGAATTCGGTGGAC 1440
   |||||
Db 1381 ATAGAAGACTCTATGCAAGGACACAGAAAGCCCATCTTTGTCAACGAATTCGGTGGAC 1440
QY 1441 GCGATAGCTGCGATCCACTACAGATCCACTCAAAATGTTCTCGAAGAGTACCAAGCAGAG 1500
   |||||
Db 1441 GCGATAGCTGCGATCCACTACAGATCCACTCAAAATGTTCTCGAAGAGTACCAAGCAGAG 1500
QY 1501 CTGCTTGAAGAAACGATCAGAGCTCTTTTGAAGAAAGACTATCATCATCGGAACACAGT 1560
   |||||
Db 1501 CTGCTTGAAGAAACGATCAGAGCTCTCTTTTGAAGAAAGACTATCATCATCGGAACACAGT 1560
QY 1561 TGGGCTTTGCAATTTTAAGCTCTCTCAAGTGTGAAGAACCAATTTCAACCAAG 1620
   |||||
Db 1561 TGGGCTTTGCAATTTTAAGCTCTCTCAAGTGTGAAGAACCAATTTCAACCAAG 1620
QY 1621 GGTGTTTTCACAAAGAGACAGACCAACCAACCTGTTGCTGATGTACTGGAAGACTGTGG 1680
   |||||
Db 1621 GGTGTTTTCACAAAGAGACAGACCAACCAACCTGTTGCTGATGTACTGGAAGACTGTGG 1680
QY 1681 AGTGAGGTT 1689
   |||||
Db 1681 AGTGAGGTT 1689
   |||||

RESULT 3
US-10-757-093-1
; Sequence 1, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757,093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patencin version 3.2
; SEQ ID NO 1
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Scopulariopsis sp. isolate RP38.3
US-10-757-093-1

Query Match 10.4%; Score 175.4; DB 22; Length 1926;
Best Local Similarity 49.0%; Pred. No. 3.2e-45;
Matches 777; Conservative 0; Mismatches 717; Indels 92; Gaps 8;
QY 172 GTTCCGAAGAACTTTCACAAAAACATCAGACTTTACTTTGCTGCGGTGAACGAGAC 231
   |||||
Db 322 GTTCCGAAGAACTTTCACAAAAACATCAGACTTTACTTTGCTGCGGTGAACGAGAG 381
QY 232 TGGCAGGTCTTCTCAACGAGAGAAAGTGGAGAGAAATCACTTGAATACCTTCCCTTC 291
   |||||
Db 382 GCCAAGGTCTTACGTCAACGAGAGAAAGTGGAGAGAAATGTGCGGTGATATATCTCCCTTC 441
   |||||

```



```
Db 1649 TTCACAGAGATCCACTGATGTTCACTGAAAGAGAACAAAGTGTGCTAGAGCAGT 1708
Qy 1513 AGCATCAGGCTCTTTTGAAGAA-----AGACTACATCATTCGGAACACAGTGTGGCCT 1567
Db 1709 ACCATCTGGGTGTGATCAAAACGACAAATATGTGTGTGAGAGCTCATTTTGAATTT 1768
Qy 1568 TTGCAGATTTTAAGACTCTGCAAAATGTGAGAAAGCCCATTCACACCAAGAGGTGTTT 1627
Db 1769 TTGCGGATTTTCACTGATGAACAGTACCGACGAGAGTGTGGGAAATAAAGGGAGTCT 1828
Qy 1628 TCACAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
Db 1829 TCACTCGGCAGACAAACCAAAAGTGCAGCTTTTTCGGA 1872

RESULT 6
US-10-272-483A-5
; Sequence 5, Application US/10272483A
; Publication No. US20040006008A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-007CP
; CURRENT APPLICATION NUMBER: US/10/272,483A
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/287,531
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 10/136,841
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 60/304,609
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/329,461
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/351,276
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant DNA sequence incorporating a signal peptide sequenc
; OTHER INFORMATION: e, the mature human beta-glucuronidase sequence, a bridge of thre
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
; OTHER INFORMATION:
US-10-272-483A-5

Query Match 10.3%; Score 174; DB 17; Length 2169;
Best Local Similarity 52.2%; Pred. No. 9,8e-45;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

Qy 737 TGGAACTTGAAGAGCGAGTACACTGTGACATCGGAATCGAATCGAATCAGTCTGGAGCG 796
Db 941 TGGGGCTGTGTCTGCTTCTTCAACACTCCCTGTGGGATCGCACTGTGCTGTACCA 1000
Qy 797 AGAAGAGGCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCAGCAGG 856
Db 1001 AGAGCAGTTCCTCATCATATGGGAAACCTTTTATTTCCACGGTGTCAACAGCATGAG 1060
Qy 857 AATTCCCGCTTGTGGGGCAGGCGACCTTTTATTCATGTATGATTAAGACTTCAACCTTC 916
```

```
Db 1061 ATGCGACATCCGAGAGGAGGCTTCGACTGGCCGCTGCTGATGAAGACTTCAACTGCG 1120
Qy 917 TGAAGTGAATCAAGCGGAATTTCTTTCAGAGACTCTCACTATCTTTACAGTGAAGGTGC 976
Db 1121 TTGCTGTGCTTGTGTCCCAAGCCTTTTCGTACAGCACTACCCCTATGCAAGAGATGA 1180
Qy 977 TGAATCTTGGCCGACAGACTCCGAAATCCCTTGTGATTAAGAAAGCCCGCAGCTGTATCA 1036
Db 1181 TGCAATGTGTGACCGCATATGGAGATGTGTGATCATGATGAATGTGCCGGCTGTGCG 1240
Qy 1037 CAAGGTACCA--CTACAATCCGAGACTCAGAAAGATAGCAAGAACAAATTAAGAA 1093
Db 1241 CGCTGCCGCAATTTCTTCAACAACGTTTCTGTGATCAACCATGTGAGTGAAGAAAG 1300
Qy 1094 TGATCGACAGACAGAAAGCAATCCCACTGATGATCATGTGAGTGTGGCGAAGCAAG 1153
Db 1301 TGTGTGTAAGAGACAAAGAACCAACCCCGGTGTGATGTGTGTGTGCGCAAGCGCTG 1360
Qy 1154 AGTCCAACTTCAGACGCGAGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db 1361 CGTCCCACTTGAATTTGCTGTGCTGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1420
Qy 1214 TGATCGAAACGCCCCGTTGTCAATGTGAGCATGATGAGACGACCAAGACAGAACAA 1273
Db 1421 TGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATGATGCAAG----- 1468
Qy 1274 GAGACGTGGCGCTGAAGTACTTCACATCGTGTGTGAACAGAGTACTAGCGGTGACA 1333
Db 1469 CAGACAAAGGGGCTCCGATGTGATGTGATCTGTTGAACAGCTACTACTTGTGTATC 1528
Qy 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTGTGAAAAAGACATAGAGACTCT 1393
Db 1529 ACGACTAGGGGACCTGAGATGTTATTCAGCTGACGAGCTGCCACCAAGTTTGAAGATG 1588
Qy 1394 ATGCAAGCACAAGAAAGCCATCTTTGTCAAGAAATTCGTGCGGACCCGATAGTGC 1453
Db 1589 ATAAGAAATATCAGAAAGCCATTTATTCAGACGAGTATGAGAGCAGAAACGATTCAGGG 1648
Qy 1454 TCCACTAGATCCACTCAATGTTCTCCGAAGATCAAGACGAGCTCGTGAAG- 1512
Db 1649 TTCACAGAGATCCACTTGAATGTCACTGAAGATCAAGAAAGTGTGCTAGAGCA 1708
Qy 1513 AGCATCAGGCTCTTTTGAAGAA-----AGACTACATCATCGGAACACAGTGTGGCCT 1567
Db 1709 ACCATCTGGGTGTGATCAAAACGACAAATATGTGTGTGAGAGCTCATTTTGAATTT 1768
Qy 1568 TTGCAGATTTTAAGACTCTCTGAGAAATGTGAGAAAGCCCATTTCTCAACCAAGAGTGT 1627
Db 1769 TTGCGGATTTTCACTGATGAACAGTACCGACGAGAGTGTGGGAAATAAAGGGAGTCT 1828
Qy 1628 TCACAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
Db 1829 TCACTCGGCAGACAAACCAAAAGTGCAGCTTTTTCGGA 1872

RESULT 7
US-10-421-175-1
; Sequence 1, Application US/10421175
; Publication No. US20030219414A1
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; CURRENT FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
```

```

; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
;
US-10-421-175-1

```

Query Match	10.2%	Score 172.4	DB 17	Length 1956
Best Local Similarity	52.1%	Pred. No. 3.1e-44		
Matches 452; Conservative	0	Mismatches 431	Indels 21	Gaps 4

QY	737	TGGAACCTTGAAAAAGACAGTACACTCTGACACTCGGAATCAAGAACGATCAAGCTGGACG	796
Db	938	TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGGATCCGACTGTGGCTGTCA	997
QY	797	AGAAAGGCTCTATCTGAAACGGAAACCTGTCTTTTGAAGGGCTTTGGAAGACGAG	856
Db	998	AGAGCAGTTCCTCATCAATGGGAACCTTTTCTATTTCCAGGATGTCAACAGCATAGG	1057
QY	857	AATTCCTCCGTCTGTGGGGCAGGGCACCTTTTATCCATTGATGATTAAGACTTCAACTTC	916
Db	1058	ATGCGGACATCCGAGGGAAAGGGCTTTCGACTGGCCGCTCTGATGAAGACTTCAACTGC	1117
QY	917	TGAAGTGATCAACGCGAATCTTTCAAGACCTCTCACTATCTTACAGTAAAGTGGC	976
Db	1118	TTCCGTGGCTTGTGTCCACACGCTTTTCGTACACGCACTTACCCTTATGCAAGAGATCGA	1177
QY	977	TGGATCTTGCCGACAGACTCGGAATCTTGTGATAGCAAGCCCCGACCGTTGGATCA	1036
Db	1178	TGCAGATGTGTGACCGCTATGGGAATGTGTGCATGATGATGTTCGCCGCTGGCCTGG	1237
QY	1037	CAAGGTACCA---CTACAAATCCGAGACTCGAAGATGACAGAAAGCAACAATTAAAGAA	1093
Db	1238	CGCTGCCGAGTTCTTCAACAACGTTTCTCTGCATCACACATGCAAGGTATGGAAGAG	1297
QY	1094	TGATCGACAGACACAMGAACCATCCAGCTGTGATCATGTGAGTGTGGCGAACAACAG	1153
Db	1298	TGTGTGCTGAGGAGACAMGAACCAACCCCGGGTGTGTGATGTGTGTGTGGCCAAACGCTG	1357
QY	1154	AGTCCAACTCATCCAGACCGCGAGGTTTCTTCAAAAGCCCTTTATGACATGCCATGAA	1213
Db	1358	CGTCCCACTTGAAATCTGTGTGGCTACATCTTGAAGATGTGATGCTTCACACCAATCTT	1417
QY	1214	TGATATCGAACACGCCCCGTTGTCATGTGTGAGCATGTAGTACGACACAGACGAGAACA	1273
Db	1418	TGGACCCCTCCCGGCTGTGACTTTTGTAGCAACTCTAACCTATGACG-----	1465
QY	1274	GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGACAGGTACTTACGGCTGTACA	1333
Db	1466	CAGACAAAGGGGGCTCCGATGTGTGATGTGATCTGTGTGAACAGCTACTTCTGTATC	1525
QY	1334	TCTATTCAGGGGAAGATTAAAGAGACTTCAAGCTCTGGAAAAAAGACTTAAGAGCTCT	1393
Db	1526	ACGACTACGGGCACTGTGATTGATTACGTCTGACGTGGCCACCAAGTTTAAAGACTGT	1585
QY	1394	ATGCAAGGCACAGAAAGCCCATCTTTGTTCACAGAAATCGTGTGGGACGCGATTACTGGCA	1453
Db	1586	ATTAAAGATTATCAGAGCCCATTTATTCAGACGAGTATGAGACGAAACGATTCAGAGGT	1645
QY	1454	TCCACTACGATCCACCTCAAAATGTTTCTCCGAAGATACACAGCAGAGCTCGTTGAAG-	1512
Db	1646	TTTCAACCAAGATCCACTTCTGTATGTTTCACTGAAGAGTACAGAAAAAGTCTGTAAAGCAGT	1705
QY	1513	ACGATTCAGGCTCTTTTGA AAAA-----AGACTACATCATGGGAACACAGTGTGGGCT	1567
Db	1706	ACCATCTGTGCTGTGATCABAAAACGCAAAATATGTGTGTGGAGAGCTCATTTTGAATT	1765
QY	1568	TTGCAGATTTTAAAGTCTCTCAGAAATGTGAGAAAGCCCATTTCTAACCAACAGGATGTT	1627
Db	1766	TTGCCGATTTTATGACTGTAAACAGTCAACCGACGAGAGTGTCTGGGGAATTTAAAGGGGATCT	1825

QY		1628 TCACAAGAGACAGACAACCCTGTTGCTCATGTACTGAGA 1671
Db		1826 TCACTCGGCAGAGACAAACAAAAGTGCAGCGTTCTTTGGCA 1869

RESULT 8
US-10-388-360-343

```

/ Sequence 343, Application US/103889360
/ Publication No. US20030225528A1
/ GENERAL INFORMATION:
/ APPLICANT: GENOMIC HEALTH
/ APPLICANT: Baker, Joffre B.
/ APPLICANT: Cronin, Maureen T.
/ APPLICANT: Kiefer, Michael C.
/ APPLICANT: Shak, Steve
/ APPLICANT: Walker, Michael Graham
/ TITLE OF INVENTION: GENE EXPRESSION PROFILING IN BIOPSIED TUMOR TISSUES
/ FILE REFERENCE: 39740-0001US
/ CURRENT APPLICATION NUMBER: US/10/388,360
/ CURRENT FILING DATE: 2003-03-12
/ PRIOR APPLICATION NUMBER: US 60/412,049
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR APPLICATION NUMBER: US 60/364,890
/ PRIOR FILING DATE: 2002-03-13
/ NUMBER OF SEQ. ID NOS: 384
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ. ID NO 343
/ LENGTH: 2191
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-388-360-343

```

Query Match	10.2%;	Score 172.4;	DB 17;	Length 2191;
Best Local Similarity	52.1%;	Pred. No. 3.3e-44;		
Matches 492; Conservative	0;	Mismatches 431;	Indels 21;	Gaps 4;

OY	73	GGAACTTTGAAAAAGACGAGTACACTCTGGACATTCGAAATCGAAGCATCGACTGGACG	796
Db	964	TGGGGCCCTGTGTCTGACTTCTACACACTCCCTGTGGGGATCGGACTGTGGCTGTCA	1023
OY	797	AGAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCACGAG	856
Db	1024	AGAGCCAGTTCCTCATCATATGGGAAACCTTTCTATTTCCACGGTGTCAACAGCATGAG	1083
OY	857	AATTCGCCGTTCTGGGGCAGGGCACCTTTATTCATTGATGATAAAGACTTCAACTTC	916
Db	1084	ATGGGACATCTCGAGGGAGGGGCTTTCGACTGGCCGCTGTGTGTAAGGACTTCAACTGC	1143
OY	917	TGAAGTGATCAACGGCAATCTTTACAGACCTCTCACTATCTTTACAGTAAGATGGC	976
Db	1144	TTCCGTGGCTTGTGGCCAAACGCTTTCGGTACAGCCACTACCCCTATGCAAGAGAAATGA	1203
OY	977	TGGAATCTTGGCGACAGACTGGGAATCCTGTGATATAGACGAAGCCCGACGTTTGTA	1036
Db	1204	TGCAGATGTGTGACCGCTATGAGATGTGGTGCATCGATGATGTTCGCGCTGGGCTGTG	1263
OY	1037	CAAGGTACCA---CTACAAATCCCGACACTGAAAGATAGCAGAAACAACATAAGAA	1093
Db	1264	CGCTGCCGAGTCTTTCAACAACGTTCTCTGTCAACACATGACAGGTATGGAAAG	1332
OY	1094	TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGGAGTGTGGCAACGAACG	1153
Db	1324	TGTGCGTAGGGAACAAGAACCAACCCGCGGTGATGTGTGTGTGGCCAACGAGCTGT	1383
OY	1154	AGTCCAAACATCCAGAGCGCGGAGGTTTCTTCAAAGCCCTTTATGAGACTGCCATGAA	1213
Db	1384	CGTCCCACTTGAATCTGTGGCTACATCTTGAAGATGTGTATGCTTCAACCAATCTCT	1443
OY	1214	TGATTCGAACAACGCCCGGTGTATGTGAGCATGATGAGACGCAACGACGAGAAACA	1273
Db	1444	TGAGACCTTCCCGGCTGTGACTTTGTGAGCAACTTAAATATGACG-----	1491
OY	1274	GAGACGTGGCGTGAAGTACTTTCACATCTGCTGTGGAACAGGTACTTACGGCTGTGACA	1333

```
Db 1492 CAGACAGGGGGCTCCGTATGTGATGTATCTGTTTGAAGCTACTACTCTTGATATC 1551
Qy 1334 TCTATCAGGGAAGATAGAAAGAGCTTCAAGCTCTGGAAAAACATAGAAAGCTCT 1393
Db 1552 AGGACTACGGGACCTCGAGTTGATTCAGCTCAGCTGGCCACCAGTTTGAAGCTGGT 1611
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTGTACAGAAATTCGGTCCGACGCGATAGCTGCA 1453
Db 1612 ATTAAGATATCAGAAAGCCATTATTCAGAGGAGATAGAGCAGAAAGATTTGACAGGT 1671
Qy 1454 TCCACTACGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAG- 1512
Db 1672 TTCACAGGATCCACTCTGATGTTTCACTGAAGAGTACAGAAAGTGTCTAGAGCAGT 1731
Qy 1513 AGGATCAGGCTCCTTTGAAAAA-----AGACTACATCTCGGAACACACGTGTGGCCT 1567
Db 1732 ACCATCTGGGCTCGATTCAGAAACGCAAAAAATGTGGTGGAGAGCTCATTTGAAAT 1791
Qy 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCACACCAAGAGGATTT 1627
Db 1792 TTGCGGATTTCTAGCTGAACAGTCCGACGAGAGTCTGGGGAATTAAGGGGATCT 1851
Qy 1628 TCACAGAGACAGACAAACCCTGTTGCTCATGTACTGAGA 1671
Db 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGGCA 1895
```

```
RESULT 9
US-10-429-802-16
; Sequence 16, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF) -RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; PRIORITY FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-16
```

```
Query Match 10.2%; Score 172.4; DB 17; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.3e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
Qy 737 TCGAATTGAAAAAGCAGTACACTCTGACATCGAATCAGAAAGATCAGCTGGAGC 796
Db 964 TGGGGCTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGCACCA 1023
Qy 797 AGAAGAGCTATCTGAAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 1024 AGAGCAGATCTCATCATATGGAAGACCTTTCTATTCCAGGTGTCAACAGATGAGG 1083
Qy 857 AATTCCCTTCTGGGGGAGGCGACTTTTATCCATTGATGATTAAGAGCTTCAACCTTC 916
Db 1084 ATGCGGACATCGAGGAGGAGGCTTGCAGTGCCTGTGTGTAAGAGCTTCAACCTTC 1143
Qy 917 TGAAGTGATCAACGGGAATCTTTTCAAGACCTCTCATATCTTTACAGTGAAGATGAC 976
Db 1144 TTGCGTGGCTTGTGTCGAACGCTTTCGATCCAGCAGCATACCCCTATGAGAGAGATGA 1203
Qy 977 TGGATCTTGGCAGACAGACTCGGAATCTTGTGATAGACGAAGCCCGCAGCTTGTGATCA 1036
Db 1204 TGCAGATGTGTACCGCTATGGGATGTGTGTCATGATGATGATCCCGGCTGGGCTGG 1263
```

```
Qy 1037 CAAGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAAGAGCAACATTAAGAGAA 1093
Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTTGCACTACCAATGAGAGGTATGAAAG 1323
Qy 1094 TGAATCAGACAGACAGAAACCATCCAGTGTATCATGTGAGTGTGGCGAAGCAACGAG 1153
Db 1324 TGTGTGTAAGGAGACAGAAACCAACCCCGGCTGTGTATGTGTGTGTGTGGCCAAAGAGCTGG 1383
Qy 1154 AGTCCACATCCAGACCGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db 1394 GGTCCACCTTAAGATCTGTGGCTAATCTTGAAGATGTATGCTCACACCAATCTCT 1443
Qy 1214 TGAATGAAACAGGCCCCGTGTGTCATGTGAGCATGATGAGACGACAGACAGAGAACAA 1273
Db 1444 TGAACCTCTCCCGGCTTGTGACTTTTGTAGCACTTAATCTATGACAG----- 1491
Qy 1274 GAGAGTGGCGCTGAGATCTTTCGACATGCTGTGTGAAACAGGTACTACGGCTGTACA 1333
Db 1492 CAGACAAAGGGGGCTCCGTATGTGATGTGATGTGATCTGTGTAACAGTACTACTTGTATC 1551
Qy 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTCTTGAAAAAGACATAGAAAGCTCT 1393
Db 1552 AGGACTACGGGCACTGAGATTGATTCAGCTGCAAGCTGGCCACCAGTTTGAAACTGGT 1611
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTGTTCACAGAAATTCGGTCCGACGCGATAGCTGCA 1453
Db 1612 ATTAAGATATCAGAAAGCCCATTTATTCAGACGAGATATGAGCAGAAACGATTTGAGGT 1671
Qy 1454 TCCACTACGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAG- 1512
Db 1672 TTCACAGGATCCACTCTGATGTTTCACTGAAGAGTACCAAGAAAGTCTGTAGAGCAGT 1731
Qy 1513 AGGATCAGGCTCCTTTGAAAAA-----AGACTACATCATGGAACACACGTGTGGCCT 1567
Db 1732 ACCATCTGGGCTCGATCAAAACGCAAAAAATGTGGTGGAGAGCTCATTTGAAAT 1791
Qy 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCACACCAAGAGGATTT 1627
Db 1792 TTGCGGATTTCTAGCTGAACAGTCCGACGAGAGTCTGGGGAATTAAGGGGATCT 1851
Qy 1628 TCACAGAGACAGACAAACCCTGTTGCTCATGTACTGAGA 1671
Db 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGGCA 1895
```

```
RESULT 10
US-10-430-503-7
; Sequence 7, Application US/10430503
; Publication No. US20040005684A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: LAN, KENG-LI
; APPLICANT: OU-YANG, FU
; APPLICANT: LIU, JAM-CHING
; APPLICANT: LAN, KENG-HSIN
; TITLE OF INVENTION: TARGETING PROTEINS TO DELIVER THERAPEUTIC OR DIAGNOSTIC
; FILE REFERENCE: UTSC:797US
; CURRENT APPLICATION NUMBER: US/10/430,503
; PRIORITY FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: 60/383,063
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-430-503-7

Query Match 10.2%; Score 172.4; DB 17; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.3e-44;
```

Matches	492;	Conservative	0;	Mismatches	431;	Indels	21;	Gaps	4;
QY	737	TGGAACTTGA	AAAAACGAGTAC	CTCTGGA	CATCGGAAT	TCGAACGAT	CACCTGGAGC	796	
Db	964	TGGGGCTGTGTCTG	CTTCTACAC	CTCCCTGTGGGGATCCG	CACTGTGGCTGTACCA	1023			
QY	797	AGAAAGGCTTATCTGA	ACGGAAAC	CTGTCTTTTGA	AGGCTTTGGAAG	CAACGAGC	856		
Db	1024	AGAGCAGTCTTCTCAT	CAATGGAAAC	CTTTCTATTTC	ACCGGTGCAAC	CAAGCATGAGG	1083		
QY	857	AATTCCTCTTGTGGG	CGAGGCA	CTTTTATCCAT	TGATGATTA	AAAGACTTCA	CTTTC	916	
Db	1084	ATGCGGACATCCGAG	GGAAGGGCTT	GACTGGCCGCTG	GTGAAGAGC	CTTCAACCTTC	1143		
QY	917	TGAATGTGATCAAC	CGGAATTTCTT	TACAGAC	CTCTCATCTAT	CTTTACAGTGAAG	GTGC	976	
Db	1144	TTTCGTGGCTGTGTG	CCAAACGCTTTC	CGTACAG	CCATACCCCTTAT	GACAGGAAGTGA	1203		
QY	977	TGGATCTTGGCGA	CGACTCGGAAT	CTTGTGTGAT	GAGAGAGCCCGG	CAAGTTGGTATCA	1036		
Db	1204	TGCAGATGTGTAC	CGCTATGGGATTTGTG	TCTCATGTGAT	GTGTCGGGCTGGG	CTGG	1263		
QY	1037	CAAGGTACCA---	CTACAAATCCCGA	ACTCAGA	AGATAGCAGAAG	CAACATAGAAGAA	1093		
Db	1264	CGCTGCGGAGTTCTT	CAACAAAGTTTCT	TGATACCA	ATGCAATGCAATGGAAG	1323			
QY	1094	TGATGACACAGACA	GAAGAACAT	TCCAGTGTGAT	CATGTGAGTGTGCG	GAAGCAACGAG	1153		
Db	1334	TGGTCGTATGGGACA	GAAGAACACCCG	CGTGTGTGTG	GTGTGCGCAACGAG	CTGC	1383		
QY	1154	AGTCCAAACATCCAG	ACGCGGAGGGTTT	GTCTCAAA	AGCCCTTTATG	AGACTGCAANTGAA	1213		
Db	1384	CGTCCACCTTAAGAT	CTGTGCTGCTACT	ATTGAAGTGTGAT	TCGCTCACACCAAT	CTCT	1443		
QY	1214	TGGATCGAACAC	AGCCCGCTTGT	CTATGTGTGAG	CATGATGAGCCAC	CAAGAGAACAA	1273		
Db	1444	TGGACCCCTCCGGG	CTGTGACCTTTGT	GTAGCAAC	CTTAATGAG	-----	1491		
QY	1274	GAGAGTGGCGCTGA	AGTACTTTCGAC	ATCGTCTGTGTGA	CAAGGTACTAG	CGCTGTGTACA	1333		
Db	1492	CAGACAAAGGGGCT	CGTATGTGTGAT	GTGATCTGTTGA	ACAGTACTACTCTT	GTGTATC	1551		
QY	1334	TCTATCAGGGAAG	ATAGAAGAGACTT	CAAGCTTGG	AAAAAGCAATAGAAG	CTCT	1393		
Db	1552	ACGACTACGGGCA	CTGTGAGTGTAT	CAGCTGCAAGTGG	CAACCGATTTGA	ACTGCT	1611		
QY	1394	ATGCAAGGACAGAA	AGCCCATCTT	TGTACAGAAAT	TCGTCGGA	CGCGATAGCTGCA	1453		
Db	1612	ATTAAGAAGTATCAG	AAGCCATTATTC	AGAGGAGAT	AGAGAAAGATTTG	CAGGGT	1671		
QY	1454	TCCACTACGATCC	ACTCAATGTTC	CCGAAGAGTAC	CAAGAGAGCTCG	TTGAAAG-	1512		
Db	1672	TTACACAGGATCC	ACTCTGATGTTC	ACTGAAGAGTAC	CAAGAAAGTGTG	CTAGAGAGT	1731		
QY	1513	ACGATCAGGCTCT	TTTGAAGAA--	---AGACTAC	ATCATTCGGA	ACACACGCTGTGGCCT	1567		
Db	1732	ACCATCTGGGTCT	GTGATCAAAA	CGCAGAAATAT	GTGTGAGAGCT	CACTTTGGAAT	1791		
QY	1568	TTGCAAGATTTT	AAGACTCTCT	CAGAAATGTG	AGAAAGACCAAT	TTCAACCAAGAGGT	1627		
Db	1792	TTGGCGATTTT	CATGACTGA	ACAGTACCG	AGAGAGTGTGG	GAATTAAGAGGGAGT	1851		
QY	1628	TCACAAGAGACAG	CAACCCCAAT	CTGTTGCT	CATGTACTGAGA	1671			
Db	1852	TCACTCGGAGAG	ACAAACCAAA	AGTGCAG	CTTCTTTTGGA	1895			

TITLE OF INVENTION: Methods for identifying marker genes for cancer									
FILE REFERENCE: 68733-A; 070/US1									
CURRENT APPLICATION NUMBER: US/10/335,053									
CURRENT FILING DATE: 2003-03-27									
PRIOR APPLICATION NUMBER: 60/345,317									
PRIORITY FILING DATE: 2001-12-31									
NUMBER OF SEQ ID NOS: 319									
SOFTWARE: PatentIn version 3.2									
SEQ ID NO 51									
LENGTH: 2191									
TYPE: DNA									
ORGANISM: Homo sapiens									
US-10-335-053-51									
Query Match									
Best Local Similarity 52.1%; Pred. No. 3.3e-44;									
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;									
QY	737	TGGAACTTGA	AAAAACGAGTAC	CTCTGGA	CATCGGAAT	TCGAACGAT	CACCTGGAGC	796	
Db	964	TGGGGCTGTGTCTG	CTTCTACAC	CTCCCTGTGGGGATCCG	CACTGTGGCTGTACCA	1023			
QY	797	AGAAAGGCTTATCTGA	ACGGAAAC	CTGTCTTTTGA	AGGCTTTGGAAG	CAACGAGC	856		
Db	1024	AGAGCAGTCTTCTCAT	CAATGGAAAC	CTTTCTATTTC	ACCGGTGCAAC	CAAGCATGAGG	1083		
QY	857	AATTCCTCTTGTGGG	CGAGGCA	CTTTTATCCAT	TGATGATTA	AAAGACTTCA	CTTTC	916	
Db	1084	ATGCGGACATCCGAG	GGAAGGGCTT	GACTGGCCGCTG	GTGAAGAGC	CTTCAACCTTC	1143		
QY	917	TGAATGTGATCAAC	CGGAATTTCTT	CAGAGAC	CTCTCATATCTT	CAAGTGAAG	GTGC	976	
Db	1144	TTTCGTGGCTGTGTG	CCAAACGCTTTC	CGTACAG	CCATACCCCTTAT	GACAGGAAGTGA	1203		
QY	977	TGGATCTTGGCGA	CGACTCGGAAT	CTTGTGTGAT	GAGAGAGCCCGG	CAAGTTGGTATCA	1036		
Db	1204	TGCAGATGTGTAC	CGCTATGGGATTTGT	GTCATGTGAT	GTGTCGGGCTGGG	CTGG	1263		
QY	1037	CAAGGTACCA---	CTACAAATCCCGA	ACTCAGA	AGATAGCAGAAG	CAACATAGAAGAA	1093		
Db	1264	CGCTGCGGAGTTCTT	CAACAAAGTTTCT	TGATACCA	ATGCAATGCAATGGAAG	1323			
QY	1094	TGATGACACAGACA	GAAGAACAT	TCCAGTGTGAT	CATGTGAGTGTGCG	GAAGCAACGAG	1153		
Db	1334	TGGTCGTATGGGACA	GAAGAACACCCG	CGTGTGTGTGAT	GTGTGCGCAACGAG	CTGC	1383		
QY	1154	AGTCCAAACATCCAG	ACGCGGAGGGTTT	GTCTCAAA	AGCCCTTTATG	AGACTGCAATGAA	1213		
Db	1384	CGTCCACCTTAAGAT	CTGTGCTGCTACT	ATTGAAGTGTGAT	TCGCTCACACCAAT	CTCT	1443		
QY	1214	TGGATCGAACAC	AGCCCGCTTGT	CTATGTGTGAG	CATGATGAGCCAC	CAAGAGAACAA	1273		
Db	1444	TGGACCCCTCCGGG	CTGTGACCTTTGT	AGAGCAAC	CTTAATGAG	-----	1491		
QY	1274	GAGAGTGGCGCTGA	AGTACTTTCGAC	ATCGTCTGTGTGA	CAAGGTACTAG	CGCTGTGTACA	1333		
Db	1492	CAGACAAAGGGGCT	CGTATGTGTGAT	GTGATCTGTTGA	ACAGTACTACTCTT	GTGTATC	1551		
QY	1334	TCTATCAGGGAAG	ATAGAAGAGACTT	CAAGCTTGG	AAAAAGCAATAGAAG	CTCT	1393		
Db	1552	ACGACTACGGGCA	CTGTGAGTGTAT	CAGCTGCAAGTGG	CAACCGATTTGA	ACTGCT	1611		
QY	1394	ATGCAAGGACAGAA	AGCCCATCTT	TGTACAGAAAT	TCGTCGGA	CGCGATAGCTGCA	1453		
Db	1612	ATTAAGAAGTATCAG	AAGCCATTATTC	AGAGGAGAT	AGAGAAAGTGTG	CTAGAGAGT	1671		
QY	1454	TCCACTACGATCC	ACTCAATGTTC	CCGAAGAGTAC	CAAGAGAGCTCG	TTGAAAG-	1512		
Db	1672	TTACACAGGATCC	ACTCTGATGTTC	ACTGAAGAGTAC	CAAGAAAGTGTG	CTAGAGAGT	1731		
QY	1513	ACGATCAGGCTCT	TTTGAAGAA--	---AGACTAC	ATCATCTGGA	ACACGCTGTGGCCT	1567		
Db	1732	ACCATCTGGGTCT	GTGATCAAAA	CGCAGAAATAT	GTGTGAGAGCT	CACTTTGGAAT	1791		


```

QY 1568 TTGCAGATTTAAGCTCTCTCAGAAATGTGAGAGAGCCCATTTCTCAACCAAGAGGTTT 1627
DB 1792 TTGGCGATTTCTAGTACTGAACAGTACCGACAGAGAGTGTGGGAAATTAAGAGGAGATCT 1851
QY 1628 TCACAGAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
DB 1852 TCAGTCGGCAGAGACAAACCAAGAGAGGCTTCTTTGGCA 1895

RESULT 12
US-10-388-934-225
; Sequence 225, Application US/10388934
; Publication No. US2004000547A1
; GENERAL INFORMATION:
; APPLICANT: Boese, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Declaf
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 225
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Rattus norvegicus (No. US2004000547A1way rat)
US-10-388-934-225

Query Match
Best Local Similarity 9.9%; Score 166.8; DB 17; Length 2472;
Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

QY 751 GACGATGACCTGTGACATCGGAATCAGAGATAGCTGGAGAGAGGCTCTAT 810
DB 946 GACTTCTAATCCCTCCCTGTGCGATTCGAACAGTGGCTGTCAACAAAGGCAAGTTCCTC 1005
QY 811 CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGACAGAGGAAATCCCGCTTCG 870
DB 1006 ATAAATGGGAACCTTTTCACTTCAAGGCGTCAACAAACAGAGATTCAGATATCGA 1065
QY 871 GGGCAGGGCACTTTTATCATGATGATTAAGAAAGACTCAACCTTGTGAAGTCAAC 930
DB 1066 GGGAGAGGCTTGACCTGCTGCTGATTAAGAAAGATTTCAACCTCCTCGTTGCTCGG 1125
QY 931 GGGAAATCTTTTGAAGCTCTGACTATCTTCAAGTGAAGAGTGGGATCTTGGCCGAC 990
DB 1126 GGAATTTCTTTGTAACAGCCATATCCCTTACTCGAGAGAGTATTCAGCTCTGTGAC 1185
QY 991 AGACTCGAATCTTGTGATAGAGAGAGCCCGCACTGTTGATATCAAGTACGA--CC 1047
DB 1186 CGATATGGAATGTGATCATCATGATGATGCCGGTGGGATGTGCTGCCAGAGT 1245
QY 1048 TACAAATCCGAGACTCAGAGATAGCAGAGACACATATAAGAAATGATGACAGACAC 1107
DB 1246 TTTGGCAAGTGTCTTCTTGGGACCACTTAAGAGTATGAGAGAGTGTGCGAGGAG 1305
QY 1108 AAGAACCATCCAGTGTGATGATGAGAGTGGGAGAGAGCAAGTCAACATGCA 1167
DB 1306 AAAAATCACCTTGGCGGTGATGATGTGCTGTGGCAATGAGCTGTCTTCTGAAA 1365
QY 1168 GACGCGAGAGTCTTTTCAAGCCCTTTATGAGATGCAATGAATGATGCAACAGC 1227
DB 1366 CCGGCGGATATTAATCTTCAAGAGCGTGAATGCCCAACCAAGCCCTGAGCCCAAGCCT 1425
QY 1228 CCGGTTGATGATGATGATGATGAGCAGCAGCAGAGAGAGAAACAAGACGTTGGCGCTG 1287
DB 1426 CCGTGTGACCTTTGTGTGAGCAATA-----CCAGATATGACGACAGACATGAGGGGCGC 1473

```

```

QY 1288 AAGTACTTGCAGATGCTGTGTGAGAGAGTACTACGGCTGTGATCATCTATCAGGAGAG 1347
DB 1474 CCGTACGTGAGAGCTGATTTGTGTGAACGTTACTTATCTGTATCATGACTCAGGGCAT 1533
QY 1348 ATAGAAGAGAGCTTCAAGCTCTGAAAAMAGACATAGAAGAGCTTATGAAAGCAGAGA 1407
DB 1534 CTGAGAGTATTCAGCTCAGCTGATGATGAGCAGAGTTTGAAGCTGGTATTAAGTATGACAG 1593
QY 1408 AAGCCATCTTTTGCAGAAATTTGGTGGGAGCCGATAGCTGGCATCATGATCA 1467
DB 1594 AAGCCAAATTAATCCAGAGGATATGAGACACCCGCTCGGGGCTTCAAGAGATCA 1653
QY 1468 CCGTCAATGTTCTCGAAGAGTACCAAGCAGAGCTCGTTGAAAAGACATCAGGCTCCTT 1527
DB 1654 CCGCATGTTTCACTGAGAGATACCAAGACAGCTCTCTCGAGAAATTAATTAATTCCTG 1713
QY 1528 TTGA-----AAAAGTACATGATGGAACACAGTGTGGCCCTTTCAGATTTTAA 1581
DB 1714 GATGAGAAAGAAAAGAAATATGTATGAGAGAGCTCATCTGAAATTTTGTGACTGATG 1773
QY 1582 ACTCTCAGATGTGAGAAAGCCATTTCTCAACCAAGAGGTGTTTCAAGAGACAGA 1641
DB 1774 ACGAACAGTCAACCACTGAGAGTAAACAGAAACAAAGAGGAGTCTTCACTGACAGAGA 1833
QY 1642 CAACCCAACTGTTGCTCATGTACTGAGA 1671
DB 1834 AACCCAAAGATGCAAGCTTCAATTTTGGCA 1863

```

```

RESULT 13
US-10-421-175-3
; Sequence 3, Application US/10421175
; Publication No. US2003021941A1
; GENERAL INFORMATION:
; APPLICANT: Podsekoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VECTORS FOR THE
; TREATMENT OF LYSOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; CURRENT FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1947)
US-10-421-175-3

```

```

Query Match
Best Local Similarity 9.4%; Score 159.6; DB 17; Length 1947;
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

QY 739 GAACCTTGAAGAAAGAGTACCTGTGACATCGGAATCAGAACGATCAGTGGAGCAG 798
DB 928 GAGTGTGACGACTACTATCAACCTTCTATCGGAGATTCGAACAGTGGCTGTCAAAAG 987
QY 799 AAGAGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAGAA 858
DB 988 AGCAAGTTCCTCATTAAGGAGAGCCCTTATTTCCAAAGGGGTCAATTAAGACAGAGAT 1047
QY 859 TTCCCGCTTGTGGGCAAGGCACTTTTATCATGATGATTAAGAACTTCAACTTCTG 918
DB 1048 TCAGATATCCGAGGAGAAAGGCTTGAATGAGCCGCTGTGTAAAGATTTCAACCTGTCTC 1107

```


Qy	919	AAGGAGATCAACGGCAATTTCTTTAGAGACCTCTCACTATCCCTTACAGTGAAGAAGTGGCTG	978
Db	1108	CGTTGGCTCGGGGCAAAATTTCTTTCTGTACACGCCACTATCCCTTATCTCAAGAGAGTACTT	1167
Qy	979	GATCTTGGCCGACAGACTCGGAATCCTTGTGTATAGACGAAGCCCGCACGTTGG--TATC	10359
Db	1168	CAGCTCTGTGACCGGATTAACGGGATTTGGTTCATTCATGATGATGTGTCGCCGTGTGGGCAATTGT	1227
Qy	1036	ACAAAGTACCACTACATTCGGGACTCAGAAAGATGACAAAGACACATTAAGAAAGATG	10959
Db	1228	CTACCTCAGAGTTTGGACACGAGTCACTTCGGCACCACTTAGAGGTATGAGAGAGCTG	1287
Qy	1096	ATCGACAGACAACGAACCATCCGAGTGTATCATGTGAGGTGGGCAAGAACCAAG	11555
Db	1288	GTTGGCCGGGACAAAATATCAACCTCGGTTGTGATGTGTCTGTGGCAATAGAGCTTCC	1347
Qy	1156	TCCAACCATCCAGACGCGGAGGGTTTCTTCAAGGCCCTTTATGAGACTGCCAATGAATG	12151
Db	1348	TCTGCTTGAAACCGCGCGCATATTACTTTAAGACGTATGATCACCACACCAAGCCCTG	1407
Qy	1216	GATGGAACAGCCCCCGTTGTCAATGTGTGACATGATGACGCACCAAGACGAGAAACAAG	1275
Db	1408	GACCTCACCCGCTCCGTGATCACTTGTGTGACA-----ACGCCAAATATGATGCA	1455
Qy	1276	GACGTGGCGCTGAAGTACTTGTGACATGCTGTGTGTGAACAGTACTACGCTGTATCATC	1335
Db	1456	GACCTGGGGGCCCGCTACGTGATGTATTTGTGTGAACACGCTACTTTTCTTGTATCAT	1515
Qy	1336	TATCAGGGAGAGATAGAAAGAGACTTCAGCTCTGAAAAAGACATAGAGAGCTTAT	1395
Db	1516	GACTATGGGCAATTGGAGGTGATTCAGCAACGCTGAATACCGATTGGAATCTGTAT	1575
Qy	1396	GCAAGGCACAGAAAGCCATCTTTGTACAGAAATTGGTGGAGCGGATAGCTGGCATC	1455
Db	1576	AAGACGCATCAGAAAGCCGATTTATCCAGAGCAGATGAGACAGACGCAATCCAGGGATC	1635
Qy	1456	CACATACGATCCACCTCAAAATGTTCTCGGAAGATGACCAAGACAGCTCGTTGAANAAGC	1515
Db	1636	CACAGAGACCCGCTCGCATGTTCAGTAGAGATGACCAAGAGCGTGTCTGGAANAATTAC	1695
Qy	1516	ATCAGGCTCTTTTGAAA-----AAAGACTATCATCTCGGAACACAGCTGGGCGCTT	1569
Db	1696	CATTCACTTCGTGATCAGAAACGTAAAGAAATACGTGTGGAGAGCTCATCTGAAATTC	1755
Qy	1570	GCAGATTTTAAAGACTCTCAGAAATGTGAAAGACCCATTTCAACCAAGAGGTGTTC	1629
Db	1756	GCCCACTTCAAGAGAACCACTACACCGCTGAGAGTAATCGGAAACAAGAAAGGATCTTC	1815
Qy	1630	ACAAAGACACAGAACCCAACTCGTGTCTCATGTACTGAGA	1671
Db	1816	ACTGCGCAGAGACGCCCAAACTTCGGCTTTATTGTCGA	1857
RESULT 14			
US-10-757-093-3			
Sequence 3, Application US/10757093			
Publication No. US20050153448A1			
GENERAL INFORMATION:			
APPLICANT: CAMBIA			
TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products			
FILE REFERENCE: 415			
CURRENT APPLICATION NUMBER: US/10/757,093			
CURRENT FILING DATE: 2004-01-14			
NUMBER OF SEQ ID NOS: 37			
SOFTWARE: PatentIn version 3.2			
SEQ ID NO 3			
LENGTH: 1905			
TYPE: DNA			
ORGANISM: Penicillium canescens isolate RPK			
US-10-757-093-3			

Query Match	8.4%;	Score 141.6;	DB 22;	Length 1905;
Best Local Similarity	49.1%;	Pred. NO. 3e-34;		

Matches 485; Conservative 0; Mismatches 449; Indels 54; Gaps 2;

Oy	744	TGAAAAGAACAAGTACCTCTGGACATCTGGAAATCAGAAACATCAGCTGGACGAGAAAG	803
Db	906	TGTAGTGAACAACCTACAAATTTGGCTACGGGGGTGCGTACTGTCAAGGTGTCCGGGTCA	965
Oy	804	GCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAAGCAGAGAAATCC	863
Db	966	ATTCTTATTAATAGGAAAGCCTTTCCTTATCCGGTTTGGCAAACTGAAAGCACAGC	1025
Oy	864	CGTCTGGGGGAGGGACACCTTTATCCATTTGATGATAAAGACTTCAACCTTGAAGT	923
Db	1026	AGTACGTGGCAAAAGACATGACCCAGCATCATGTCTCACATTTCCAACTCATGAATG	1085
Oy	924	GATCAACGCCAATCTTTTCAGAACCTCTCATCTATCTTACAGTGAAGGTGGCTGATCT	983
Db	1086	GATTGGAGCAAAATCTTTTGGGACTTTCACACTATCTTTACCGGGAAAGGTCAATGATTT	1145
Oy	984	TGCCGACAGACTCGGAAATCCTTGTGTAGACGAAGCCCGCAGCTGGT-----	1032
Db	1146	CGCAGATCGAAATGGAATGTGTCGTATCGATGAAACACTGCGCTGGTGTGAACATTCG	1205
Oy	1033	-----ATCAAGAAGTACCATCAATCC	1055
Db	1206	CTTGATGGGGGTATCTGAGAGTGTGCCCAAAACATTTACGCCAGATGCGATTAACGA	1265
Oy	1056	CGAGACTCAGAAAGTACAGAAAGACAATTAAGAAATGATGCACAGACACAAGAAACA	1115
Db	1266	TAAACCCAGAGGCCCAACAAGAGGGATTCGTGACTCATTTGCCAGAGCAAAAACA	1325
Oy	1116	TCCCACTGTGATCATGTGAGAGTGTGGCAAGAAACCAAGATCCAAACATCCAGACGGGA	1175
Db	1326	TGCCAGGTGTGTCAATGTGTCTATTTGGCAACGAGCCCGCATCTCATGAAGATGAGGTG	1385
Oy	1176	GGGTTCTTCAAAAGCCCTTTATGAGACTGSCAAATGAATGATGGAACAGCCCGTGT	1235
Db	1386	CGAATCTTGGAGCCACTGACCAATTTGACTTCGTCACTTGATCAACTGCGCTTATAC	1445
Oy	1236	CATGCTAGCATGTAGACGCAACGACGAGAGAACAGAGACGTGGCGCTGAAGTACTT	1295
Db	1446	ATTGTCTAAGTCGGGACGGGACATATCAAGCTGGATCGGATCTGTATCTG-----TT	1499
Oy	1296	CGACATGCTCTGTGTAAACAGTACTACGCTGTGATCATCTATCAGGAAGAGATGAGA	1355
Db	1500	TGAATGAGTTCATAATCGGTATTTCCGATGGTATTTCCAAACAGAGACCTTGAAGA	1559
Oy	1356	AGACCTTCAAGCTCTGGAAAAAGCATAGAAGAGCTTATGCAAGGACAGAAAGCCAT	1415
Db	1560	AGCAGAGGCAAGCTTTTAAAGAGAGCTGATGATGGCAAGAGAAATTCACAGGGCAT	1619
Oy	1416	CTTGTTCACAGAAATTCGGTCCGACGCGATAGCTGTGCATCCACTACGACTCTCAAT	1475
Db	1620	CGTCATACCGGAATATGTGCAAGTACCCTTGGCAGGCTTCACTATCTCTGGACTGCC	1679
Oy	1476	GTTTCTCCGAAGATCACAGAGAGCTGTTGAAAGACATCAGGCTCCTTTGAAAAA	1535
Db	1680	TTGAGAGGAAGAGTTCCAAATGCTAAGCATGTACATCAATCAAGTGTGTGATCCAT	1739
Oy	1536	AGACTACATCATGCGGAACACAGGTGGGCTTTGCGATTTTAAAGCTCCTCAGATGT	1595
Db	1740	TGAATCGATGACAGGCGAGCATGTTTGGAACTTCGCGAATTTCCAGACCAACTTGGGAT	1799
Oy	1596	GAGAAAGCCCATTTCTCAACCAACAAGGTGTTTTTCACAAGAGACGACAAACCAACTCGT	1655
Db	1800	CATCCGAGTGAAGGTATCAAGAAAGGGTGTTCACCCGTGACCGAAAGCCAAAGCGGC	1859
Oy	1656	TGCTCATGTACTGAAGAAGCTGTGGAGT	1683
Db	1860	AGCTCATAGTTTGAAGGCAAGGTGGACT	1887

RESULT 15
US-10-120-145-7

```

; Sequence 7, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Killian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1887
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-10-120-145-7

```

```

Query Match      8.4%; Score 141.4; DB 16; Length 1887;
Best Local Similarity 45.4%; Pred. No. 3.5e-34;
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

```

```

QY 120 CCAAGATCTGTCTCGAAGAAAGACCCCTTCACCTCAAAACACCTTCTACCTCCGA 179
DB 208 CAAGAAATCCCGCAACCAATATGCTGTGTAACGATCGTACCGTCCGAC 267
QY 180 GNAACCTTCAAAACACATCAGACTTACTTGTGCGGTGAACACGAGCTGGAGAT 239
DB 268 CTAATCGAAGATCAGCCGATGCTGCTCCGCTTCGCTGCACTCAAAAGCAATGT 327
QY 240 CTTCTCTCAACGAGAAAGAGAGAGAGATCAATGATACCTTCCCTTGAAGTGA 299
DB 328 CTATGCAATGTGAGCTGTGCTGTGAGCAAAAGGCGGATTCCTGCACTTGAAGCGGA 387
QY 300 TGTGACGGGAAAGTGAATTCGGAGAGAACGAACTCAGGAGTGTGTTGAGAACAGAT 359
DB 388 AATCAACAACCTGCTGCGATGATGGAATGCGCTCACCGTGGCGGTGAGAACATCT 447
QY 360 GAAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGATTTT 419
DB 448 GCAAGATAGCACCTCCCGGTGGGCTGTACAGCCAGCGCCACGAAAGGCGCTCGGAAA 507
QY 420 TGGAAATTTCCACCTGCAAACTTGCATTTTCCCTTCCCTTCCGTTGAATCAAGGCTGT 479
DB 508 AGTCAATTCGTAACAAGCCGAACTTGCATTTTCAACTATGCAAGGCTGTGCAACGTCGGT 567
QY 480 TCTGATAGAGTTCAACAGACCAAGGATTAATCTGACATCTGGGTGAGACAGAGTGAATC 539
DB 568 GAAATATCTACACGACCCCGTTTACGTACGTGAGGACATCTGGTGTGACCGACTTCA 627
QY 540 TGAACCGGAGAGAACTTGAAGAAAGTGAAGATGAAGTCTCGAAGAGCGGT 599
DB 628 TGGCCCAACCGGAGCTGTACCTATACGTGAGCTTTCAAGGCAAAAGCCGAGACCTGTGA 687
QY 600 GGGACAGAGATGACGATCAAACTTGGAGAGAGAGAGAAAGATTAGAACATCCAAAG 659
DB 688 AGTGTGCGGTGTGATGAGAGAGGCAAGTGTGCGAAGCACCGAGGGCTGAGGGTGA 747
QY 660 ATTCTGCAAGAGGAGTTCATCTCGAAGAAAGCGCAGGTTCGAGGCTTGAAGATTCATA 719
DB 748 CGTGAAGATTCGAAATGTATCTCTGGGAACCACTGAACAGTATCTCTACATCA 807
QY 720 TCTTATCTCTCAAGTGAAGTGAAGTGAAGAGAGATCACTTGAAGATCGAATCAG 779
DB 808 AGTGAAGCTGTGAGAGAGAGAGATCACTGATGTCTATGAAGAGCGGTTCGGCGTGG 867
QY 780 AAGATCAAGCTGGAGAGAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGG 839

```

```

DB 868 GACCGTGAAGTCAACGACGCGCAAGTTCCTCATCAACAAACCGTTCTACTTCAAGG 927
QY 840 CTTTGGAAAGCAGAGGAATTCCTCCGTTCTGGGGCAGGGGACCTTTATCCATGATAT 899
DB 928 CTTTGGCAAAATAGAGGACACTCTTATCAACGGCGGTGCTTTTAAAGAGGAGCAATGT 987
QY 900 AAAAGACTTCAACCTTTCGAGTGAATCAACGCAATTTCTTTCAGAGCCTTCACTATCC 959
DB 988 GATGATTTCAATATCTCTCAAAATGATGCGGCCCAACAGCTTCCGACCGCACTATCC 1047
QY 960 TTACAGTGAAGAGTGGCTGTATCTTTCGCAAGACTCGGAATCTTGTATGAGCAAGC 1019
DB 1048 GTACTTGAAGAGTGTATGCTCTTTCGGAGATCGGAGGCTGTGATGATGAGAGAC 1107
QY 1020 CCGGCACTTGTAT----- 1034
DB 1108 TCCGCACTTGGCGTGCACCTCAACTTCATGAGCAACGAGGACTCGGAGAGGAGCA 1167
QY 1035 ---CAAGATACCACTCAATCCGAGACTCAGAAAGTACAGAAAGCAATTAAGAG 1091
DB 1168 GCGGCTCATGACTCGGAGAAAGATTCGAGCCTTGAAGACATCAAGACGTTCTCCGTGA 1227
QY 1092 AATGATCAGACAGACAAAGAACATCCCATGTGTATCATGTGAGTGTGGCAAGAAC 1151
DB 1228 ACTGGTGTCTGTGAACAAAGAACATCCAAAGCGTGTGTATGTGAGCATCGCAACGAGG 1287
QY 1152 AGATTCCAACATCAGACGCGGAGGTTTCTTAAAGCCTTTATGAGACTCGCAATGA 1211
DB 1288 GGCACATGAGGAAGAGGCGGTGACGATCTTCAAGCGTGTGTGAGCTGACCAAGGA 1347
QY 1212 AATGATGCA---ACAGCCCGGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1268
DB 1348 ACTGACCCACAGAAAGGCTCCGCTACAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1407
QY 1269 AACAAGACGTCGCGCTGAAGTACTTGCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1328
DB 1408 GGAACAAAGTCCGCAATGCA-----TTGAGTGTATCGCGCTCAATCGCTTAAACGAGATG 1461
QY 1329 GTACATCTATCAGGAAAGATTAAGAAAG---GACTTCAACTCTGGAAGAAAGACATGA 1385
DB 1462 GTACTTGTATGAGGAGTATCTGAAAGGCGCCAAAGTCCATCTCGCCAGGAATTTCAAGC 1521
QY 1386 AGACTCTATGCAAGGACAGAAAGCCATCTTGTGACAGAAATTCGGTGGGACGCGAT 1445
DB 1522 GTGGAACAAAGGTTGCCAGAAAGCCGATCATATCATCTGATGATGATGATGATGATGATGAT 1581
QY 1446 AGTGGCATCCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1505
DB 1582 TGGCGGCTTTGACAGCAATGATTCAGATGATGATTCACGAGGAAATATCAAGTGTGATCTA 1641
QY 1506 TGAAGAGCATCAAGGCTCTTTTGAAGAAAGACATCAATCTGGAACACACGATGTGGG 1565
DB 1642 CAGGCGAAACAGCTCGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1701
QY 1566 CTTTGCAGATTTTAAAGCTCTCAGAAATGTGAGAAAGCCATTTCTCAACCAACCAAGGATGT 1625
DB 1702 CTTGCGGACTTCCGACCTTCAAGGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1761
QY 1626 TTTCAAGAGACAGACAAACCAATCTGTTGTATGATGATGATGATGATGATGATGATGATGAT 1681
DB 1762 GTTCACTGTGACCGCAAGCGCAAGCTGCGGCCACGCTTGTGCGAGCGCTGGA 1817

```

Search completed: September 1, 2005, 21:26:50
Job time : 1170 secs